electrophoresis (lanes 3 and 4). Of significance in lane 4 is the retention of the FLAG epitope indicating the formation of a disulfide bond between the cysteine in the CF pre sequence with a cysteine in the catalytic domain of prostasin which is presumably Cys-122 (chymotrypsin numbering). Retention of the FLAG epitope, following EK cleavage and denaturation without DTT, is not observed using the prolactin pre sequence which lacks a cysteine residue (Compare lane 4 of Figure 7 with lane 4 of Figure 8). This documents that the CF pre sequence is capable of forming a light chain, that is disulfide bonded to the heavy catalytic chain of the recombinant serine proteases, when expressed in this system. It appears that in the absence of the reducing agent DTT, the EK cleaved polypeptides have a reproducibly decreased mobility in the gel (compare lane B3 with B4).

Figure 9 - Polyacrylamide gel and Western blot analyses of the recombinant protease PFEK1-neuropsin-6XHIS expressed, purified and activated from the activation construct of SEQ.ID.NO.:9 (Figure 5). Shown is the polyacrylamide gel containing samples of the serine protease PFEK1-neuropsin-6XHIS stained with Coomassie Brilliant Blue (A). The relative molecular masses are indicated by the positions of protein standards (M). In the indicated lanes, the purified zymogen was either untreated (-) or digested with EK (+) which was used to cleave and activate the zymogen into its active form. A Western blot of the gel in A, probed with the anti-FLAG MoAb M2, is also shown. This demonstrates the quantitative cleavage of the expressed and purified zymogen to generate the processed and activated protease. Since the FLAG epitope is located just upstream of the of the EK1 pro sequence, cleavage with EK1 generates a FLAG-containing polypeptide which is too small to be retained in the polyacrylamide gel, and is therefore not detected in the +EK lane.

Figure 10 - Polyacrylamide gel and Western blot analyses of the recombinant protease PFEK1-protease O-6XHIS expressed, purified and activated from the activation construct of SEQ.ID.NO.:10 (Figure 6). Shown is the polyacrylamide gel containing samples of the novel serine protease PFEK1-protease O-6XHIS stained with Coomassie Brilliant Blue (A). The relative molecular masses are indicated by the positions of protein standards (M). In the indicated lanes, the purified zymogen was either untreated (-) or digested with EK (+) which was used to cleave and activate the zymogen into its active

Topical application of the compounds or modulators is possible through the use of a liquid drench or a shampoo containing the instant compounds or modulators as an aqueous solution or suspension. These formulations generally contain a suspending agent such as bentonite and normally will also contain an antifoaming agent. Formulations containing from 0.005 to 10% by weight of the active ingredient are acceptable. Preferred formulations are those containing from 0.01 to 5% by weight of the instant compounds or modulators.

Proteases are used in non-natural environments for various commercial purposes including laundry detergents, food processing, fabric processing, and skin care products. In laundry detergents, the protease is employed to break down organic, poorly soluble compounds to more soluble forms that can be more easily dissolved in detergent and water. In this capacity the protease acts as a "stain remover." Examples of food processing include tenderizing meats and producing cheese. Proteases are used in fabric processing, for example, to treat wool in order prevent fabric shrinkage. Proteases may be included in skin care products to remove scales on the skin surface that build up due to an imbalance in the rate of desquamation. Common proteases used in some of these applications are derived from prokaryotic or eukaryotic cells that are easily grown for industrial manufacture of their enzymes, for example a common species used is *Bacillus* as described in United States patent 5,217,878. Alternatively, United States Patent 5,278,062 describes serine proteases isolated from a fungus, *Tritirachium album*, for use in laundry detergent compositions. Unfortunately use of some proteases is limited by their potential to cause allergic reactions in sensitive individuals or by reduced efficiency when used in a non-natural environment. It is anticipated that protease proteins derived from non-human sources would be more likely to induce an immune response in a sensitive individual. Because of these limitations, there is a need for alternative proteases that are less immunogenic to sensitive individuals and/or provides efficient proteolytic activity in a non-natural environment. The advent of recombinant technology allows expression of any species' proteins in a host suitable for industrial manufacture.

Another aspect of the present invention relates to compositions comprising the Protease MH2, F, prostasin, O, and neuropsin or any other protease and an acceptable carrier. The composition may be any variety of compositions that requires a protease component. Particularly preferred are compositions that may come in contact with humans, for example, through use or manufacture. The use of the Protease MH2, F, prostasin, O, and neuropsin or any other protease of the present invention is believed to reduce or eliminate the immunogenic response users and/or handlers might otherwise experience with a similar composition containing a known protease, particularly a protease of non-human origin. Preferred compositions are skin care compositions and laundry detergent compositions.

Herein, "acceptable carries" includes, but is not limited to, cosmetically-acceptable carriers, pharmaceutically-acceptable carriers, and carriers acceptable for use in cleaning compositions.

Skin Care Compositions

Skin care compositions of the present invention preferably comprise, in addition to the Protease MH2, F, prostasin, O, and neuropsin or any other protease, a cosmetically- or pharmaceutically-acceptable carrier.

Herein, "cosmetically-acceptable carrier" means one or more compatible solid or liquid filler diluents or encapsulating substances which are suitable for use in contact with the skin of humans and lower animals without undue toxicity, incompatibility, instability, irritation, allergic response, and the like, commensurate with a reasonable benefit/risk ratio.

Herein, "pharmaceutically-acceptable" means one or more compatible drugs, medicaments or inert ingredients which are suitable for use in contact with the tissues of humans and lower animals without undue toxicity, incompatibility, instability, irritation, allergic response, and the like, commensurate with a reasonable. benefit/risk ratio. Pharmaceutically-acceptable carriers must, of course, be of sufficiently high purity and

sufficiently low toxicity to render them suitable for administration to the mammal being treated.

Herein, "compatible" means that the components of the cosmetic or pharmaceutical compositions are capable of being commingled with the Protease MH2, F, prostasin, O, and neuropsin or any other protease, and with each other, in a manner such that there is no interaction which would substantially reduce the cosmetic or pharmaceutical efficacy of the composition under ordinary use situations.

Preferably the skin care compositions of the present invention are topical compositions, i.e., they are applied topically by the direct laying on or spreading of the composition on skin. Preferably such topical compositions comprise a cosmetically- or pharmaceutically acceptable topical carrier.

The topical composition may be made into a wide variety of product types. These include, but are not limited to, lotions, creams, beach oils, gels, sticks, sprays, ointments, pastes, mousses, and cosmetics; hair care compositions such as shampoos and conditioners (for, e.g., treating/preventing dandruff); and personal cleansing compositions. These product types may comprise several carrier systems including, but not limited to, solutions, emulsions, gels and solids.

Preferably the carrier is a cosmetically or pharmaceutically acceptable aqueous or organic solvent. Water is a preferred solvent. Examples of suitable organic solvents include: propylene glycol, polyethylene glycol (200-600), polypropylene glycol (425-2025), propylene glycol-14 butyl ether, glycerol, 1,2,4butanetriol, sorbitol esters, 1,2,6-hexanetriol, ethanol, isopropanol, butanediol, and mixtures thereof. Such solutions useful in the present invention preferably contain from about 0.001% to about 25% of the Protease MH2, F, prostasin, O, and neuropsin or any other protease, more preferably from about 0.1% to about 10% more preferably from about 0.5% to about 5%; and preferably from about 50% to about 99.99% of an acceptable aqueous or organic solvent, more preferably from about 90% to about 99%.

Skin care compositions of the present invention may further include a wide variety of additional oil-soluble materials and/or water-soluble materials conventionally used in

topical compositions, at their art-established levels. Such additional components include, but are not limited to: thickeners, pigments, fragrances, humectants, proteins and polypeptides, preservatives, pacifiers, penetration enhancing agents, collagen, hyaluronic acid, elastin, hydrolysates, primrose oil, jojoba oil, epidermal growth factor, soybean saponins, mucopolysaccharides, Vitamin A and derivatives thereof, Vitamin B2, biotin, pantothenic acid, Vitamin D, and mixtures thereof.

Cleaning Compositions

Cleaning compositions of the present invention preferably comprise, in addition to the Protease MH2, F, prostasin, O, and neuropsin or any other protease, a surfactant. The cleaning composition may be in a wide variety of forms, including, but not limited to, hard surface cleaning compositions, dish-care cleaning compositions, and laundry detergent compositions.

Preferred cleaning compositions are laundry detergent compositions. Such laundry detergent compositions include, but not limited to, granular, liquid and bar compositions. Preferably, the laundry detergent composition further comprises a builder.

The laundry detergent composition of the present invention contains the Protease MH2, F, prostasin, O, and neuropsin or any other protease at a level sufficient to provide a "cleaning-effective amount". The term "cleaning effective amount" refers to any amount capable of producing a cleaning, stain removal, soil removal, whitening, deodorizing, or freshness improving effect on substrates such as fabrics, dishware and the like. In practical terms for current commercial preparations, typical amounts are up to about 5 mg by weight, more typically 0.01 mg to 3 mg, of active enzyme per gram of the detergent composition. Stated another way, the laundry detergent compositions herein will typically comprise from 0.001% to 5%, preferably 0.01%-3%, more preferably 0.01% to 1% by weight of raw Protease MH2, F, prostasin, O, and neuropsin or any other protease preparation. Herein, "raw Protease MH2, F, prostasin, O, and neuropsin or any other protease preparation" refers to preparations or compositions in which the Protease MH2, F, prostasin, O, and neuropsin or any other protease is contained in prior to its addition to

the laundry detergent composition. Preferably, the Protease MH2, F, prostasin, O, and neuropsin or any other protease is present in such raw Protease MH2, F, prostasin, O, and neuropsin or any other protease preparations at levels sufficient to provide from 0.005 to 0.1 Anson units (AU) of activity per gram of raw Protease MH2, F, prostasin, O, and neuropsin or any other protease preparation. For certain detergents, such as in automatic dishwashing, it maybe desirable to increase the active Protease MH2, F, prostasin, O, and neuropsin or any other protease content of the raw Protease MH2, F, prostasin, O, and neuropsin or any other protease preparation in order to minimize the total amount of non-catalytically active materials and thereby improve spotting/filming or other end-results. Higher active levels may also be desirable in highly concentrated detergent formulations.

Preferably, the laundry detergent compositions of the present invention, including but not limited to liquid compositions, may comprise from about 0.001% to about 10%, preferably from about 0.005% to about 8%, most preferably from about 0.01% to about 6%, by weight of an enzyme stabilizing system. The enzyme stabilizing system can be any stabilizing system that is compatible with the Protease MH2, F, prostasin, O, and neuropsin or any other protease, or any other additional detersive enzymes that may be included in the composition. Such a system may be inherently provided by other formulation actives, or be added separately, e.g., by the formulator or by a manufacturer of detergent-ready enzymes. Such stabilizing systems can, for example, comprise calcium ion, boric acid, propylene glycol, short chain carboxylic acids, boronic acids, and mixtures thereof, and are designed to address different stabilization problems depending on the type and physical form of the detergent composition.

The detergent composition also comprises a detersive surfactant. Preferably the detergent composition comprises at least about 0.01% of a detersive surfactant; more preferably at least about 0.1%; more preferably at least about 1 %; more preferably still, from about 1 % to about 55%.

Preferred detersive surfactants are cationic, anionic, nonionic, ampholytic, zwitterionic, and mixtures thereof, further described herein below. Non-limiting examples of detersive surfactants useful in the detergent composition include, the conventional C11-

C18 alkyl benzene sulfonates ("LAS") and primary, branched-chain and random C10-C20 alkyl sulfates ("AS"), the C10-C18 secondary (2,3) alkyl sulfates of the formula $CH_3(CH_2)x(CHOSO_3\text{-}M+)\ CH_3$ and $CH_3\ (CH_2)_y(CHOSO_3\text{-}M+)\ CH_2CH_3$ where x and (y + 1) are integers of at least about 7, preferably at least about 9, and M is a water-solubilizing cation, especially sodium, unsaturated sulfates such as oleyl sulfate, the C10-C18 alkyl alkoxy sulfates ("AExS"; especially EO 1-7 ethoxy sulfates), C10-C18 alkyl alkoxy carboxylates (especially the EO 1-5 ethoxycarboxylates), the C10-18 glycerol ethers, the C10-C18 alkyl polyglycosides and their corresponding sulfated polyglycosides, and C12-C18 alpha-sulfonated fatty acid esters. If desired, the conventional nonionic and amphoteric surfactants such as the C12-C18 alkyl ethoxylates ("AE") including the so-called narrow peaked alkyl Ethoxylates and C6-C12 alkyl phenol alkoxylates (especially ethoxylates and mixed ethoxy/propoxy), C12-C18 betaines and solfobetaines ("sultaines"), C10-C18 amine oxides, and the like, can also be included in the overall compositions. The C10-C18 N-alkyl polyhydroxy fatty acid amides can also be used. Typical examples include the C12-C18 N-methylglucamides. See WO 9,206,154. Other sugar-derived surfactants include the N-alkoxy polyhydroxy fatty acid amides, such as C10-C18 N-(3-methoxypropyl) glucamide. The N-propyl through N-hexyl C12-C18 glucamides can be used for low sudsing. C10-C20 conventional soaps may also be used. If high sudsing is desired, the branched-chain C10-C16 soaps may be used. Mixtures of anionic and nonionic surfactants are especially useful. Other conventional useful surfactants are listed in standard texts.

Detergent builders are also included in the laundry detergent composition to assist in controlling mineral hardness. Inorganic as well as organic builders can be used. Builders are typically used in fabric laundering compositions to assist in the removal of particulate soils.

The level of builder can vary widely depending upon the end use of the composition and its desired physical form. When present, the compositions will typically comprise at least about 1% builder. Liquid formulations typically comprise from about 5% to about 50%, more typically about 5% to about 30%, by weight, of detergent builder.

Granular formulations typically comprise from about 10% to about 80%, more typically from about 15% to about 50% by weight, of the detergent builder. Lower or higher levels of builder, however, are not excluded.

Inorganic or P-containing detergent builders include, but are not limited to, the alkali metal, ammonium and alkanolammonium salts of polyphosphates (exemplified by the tripolyphosphates, pyrophosphates, and glassy polymeric meta-phosphates), phosphonates, phytic acid, silicates, carbonates (including bicarbonates and sesquicarbonates), sulphates, and aluminosilicates. However, non-phosphate builders are required in some locales. Importantly, the compositions herein function surprisingly well even in the presence of the so-called "weak" builders (as compared with phosphates) such as citrate, or in the so-called "underbuilt' situation that may occur with zeolite or layered silicate builders.

Examples of silicate builders are the alkali metal silicates, particularly those having a SiO2:Na2O ration in the range 1.6:1 to 3.2:1 and layered silicates, such as the layered sodium silicates described in U.S. Patent 4,664,839, issued May 12, 1987 to H. P. Rieck. NaSKS-6 is the trademark for a crystalline layered silicate marketed by Hoechst (commonly abbreviated herein as "SKS-6"). Unlike zeolite builders, the Na SKS-6 silicate builder does not contain aluminum. NaSKS-6 has the delta-Na2SiO5 morphology form of layered silicate. It can be prepared by methods such as those described in German DE-A-3,417,649 and DE-A-3,742,043. SKS-6 is a highly preferred layered silicate for use herein, but other such layered silicates, such as those having the general formula NaMSixO2x+l yH20 wherein M is sodium or hydrogen, x is a number from 1.9 to 4, preferably 2, and y is a number from 0 to 20, preferably 0 can be used herein. Various other layered silicates from Hoechst include NaSKS-5, NaSKS-7 and NaSKS-1 1, as the alpha, beta and gamma forms. As noted above, the delta-Na2SiO5 (NaSKS-6 form) is most preferred for use herein. Other silicates may also be useful such as for example magnesium silicate, which can serve as a crispening agent in granular formulations, as a stabilizing agent for oxygen bleaches, and as a component of suds control systems.

Examples of carbonate builders are the alkaline earth and alkali metal carbonates as disclosed in German Patent Application No. 2,321,001 published on November 15, 1973.

Aluminosilicate builders are useful in the present invention. Aluminosilicate builders are of great importance in most currently marketed heavy duty granular detergent compositions, and can also be a significant builder ingredient in liquid detergent formulations. Aluminosilicate builders include those having the empirical formula:

$$M_z(zAlO_2)_y\text{-}xH_2O$$

wherein z and y are integers of at least 6, the molar ratio of z to y is in the range from 1.0 to about 0.5, and x is an integer from about 15 to about 264.

Useful aluminosilicate ion exchange materials are commercially available. These aluminosilicates can be crystalline or amorphous in structure and can be naturally-occurring aluminosilicates or synthetically derived. A method for producing aluminosilicate ion exchange materials is disclosed in U.S. Patent 3,985,669, Krummel, et al, issued October 12, 1976. Preferred synthetic crystalline aluminosilicate ion exchange materials useful herein are available under the designations Zeolite A, Zeolite P (b), Zeolite MAP and Zeolite X. In an especially preferred embodiment, the crystalline aluminosilicate ion exchange material has the formula:

$$Na_{12}[(AlO_2)_{12}(SiO_2)_{12}].xH_2O$$

wherein x is from about 20 to about 30, especially about 27. This material is known as Zeolite A. Dehydrated zeolites (x = 0 - 1 0) may also be used herein. Preferably, the aluminosilicate has a particle size of about 0.1-10 microns in diameter.

Organic detergent builders suitable for the purposes of the present invention include, but are not restricted to, a wide variety of polycarboxylate compounds. As used herein, "polycarboxylate" refers to compounds having a plurality of carboxylate groups, preferably at least 3 carboxylates. Polycarboxylate builder can generally be added to the composition in acid form, but can also be added in the form of a neutralized salt. When utilized in salt form, alkali metals, such as sodium, potassium, and lithium, or alkanolammonium salts are preferred.

Included among the polycarboxylate builders are a variety of categories of useful materials. One important category of poiycarboxylate builders encompasses the ether polycarboxylates, including oxydisuccinate, as disclosed in Berg, U.S. Patent 3,128,287, issued April 7, 1964, and Lamberti et al., U.S. Patent 3,635,830, issued January 18, 1972. See also "TMSFTDS" builders of U.S. Patent 4,663,071, issued to Bush et al., on May 5, 1987. Suitable ether polycarboxylates also include cyclic compounds, particularly alicyclic compounds, such as those described in U.S. Patents 3,923,679 to Rapko, issued December 2,, 1975; 3,835,163 to Rapko, issued September 10, 1974; 4,158,635 to Crutchfield et al., issued June 19, 1979; 4,120,874 to Crutchfield et al., issued October 17, 1978; and 4,102,903 to Crutchfield et al., issued July 25, 1978.

Other useful detergency builders include the ether hydroxypolycarboxylates, copolymers of maleic anhydride with ethylene or vinyl methyl ether, 1, 3,, 5-trihydroxy benzene-2, 4, 6-t6sulphonic acid, and carboxymethyloxysuccinic acid, the various alkali metal, ammonium and substituted ammonium salts of polyacetic acids such as. ethylenediamine tetraacetic acid and nitrilotriacetic acid, as well as polycarboxylates such as Mellitic acid, succinic acid, oxydisuccinic acid, polymaleic acid, benzene 1,3,5-tricarboxylic acid, carboxymethyloxysuccinic acid, and soluble salts thereof,

Citrate builders, e.g., citric acid and soluble salts thereof (particularly sodium salt), are polycarboxylate builders of particular importance for heavy-duty liquid detergent formulations due to their availability from renewable resources and their biodegradability. Citrates can also be used in granular compositions, especially in combination with zeolite and/or layered silicate builders. Oxydisuccinates are also especially useful in such compositions and combinations.

Also suitable in the detergent compositions of the present invention are the 3,3-dicarboxy-4-oxa-1,6-hexanedioates and the related compounds disclosed in U.S. Patent 4,566,984 to Bush, issued January 28, 1986. Useful succinic acid builders include the C5-C20 alkyl and alkenyl succinic acids and salts thereof. A particularly preferred compound of this type is dodecenylsuccinic acid. Specific examples of succinate builders include: laurylsuccinate, myristylsuccinate, paimitylsuccinate, 2-dodecenylsuccinate (preferred),

2pentadecenylsuccinate, and the like. Lauryisuccinates are the preferred builders of this group, and are described in European Patent Application 200,263 to Barrat et al., published November 5, 1986.

Other suitable polycarboxylates are disclosed in U.S. Patent 4,144,226, Crutchfield et al, issued March 13, 1979 and in U.S. Patent 3,308,067, Diehl, issued March 7, 1967. See also U.S. Patent 3,723,322 to Diehl, issued March 27, 1973.

Fatty acids, e.g., C12-C18 monocarboxylic acids, can also be incorporated into the compositions alone, or in combination with the aforesaid builders, especially citrate and/or the succinate builders, to provide additional builder activity. Such use of fatty acids will generally result in a diminution of sudsing, which should be taken into account by the formulator.

In situations where phosphorus-based builders can be used, and especially in the formulation of bars used for hand-laundering operations, the various alkali metal phosphates such as the well-known sodium tripolyphosphates, sodium pyrophosphate and sodium orthophosphate can be used. Phosphonate builders such as ethane-l-hydroxy-l,l-diphosphonate and other known phosphonates (see, for example, U.S. Patents 3,159,581 to Diehl, issued December 1, 1964; 3,213,030 to Diehl, issued October 19, 1965; 3,400,148 to Quimby, issued September 3, 1968; 3,422,021 to Roy, issued January 14, 1969; and 3,422,137 to Quimby, issued January 4, 1969) can also be used.

Additional components which may be used in the laundry detergent compositions of the present invention include, but are not limited to: alkoxylated polycarboxylates (to provide, e.g., additional grease stain removal performance), bleaching agents, bleach activators, bleach catalysts, brighteners, chelating agents, clay soil removal / anti-redeposition agents, dye transfer inhibiting agents, additional enzymes (including lipases, amylases, hydrolases, and other proteases), fabric softeners, polymeric soil release agents, polymeric dispersing agents, and suds suppressors.

The compositions herein may further include one or more other detergent adjunct materials or other materials for assisting or enhancing cleaning performance, treatment of

Plasmid manipulations:

All molecular biological methods were in accordance with those previously described (Sambrook, et al. *Molecular Cloning: A Laboratory Manual*, 2nd ed., (1989). 1-1626). Oligonucleotides were purchased from Ransom Hill Biosciences (Ransom Hill, CA)(Table 1) and all restriction endonucleases and other DNA modifying enzymes were from New England Biolabs (Beverly, MA) unless otherwise specified. Constructs were initially made in the pCDNA3 (InVitrogen, San Diego, CA) or the pCIneo (Promega, Madison. WI) vectors and subsequently transferred into Drosophila expression vectors pRM63 and pFLEX64 as described below. The Drosophila expression vectors used are similar to those commercially available (InVitrogen, San Diego, CA). All construct manipulations were confirmed by dye terminator cycle sequencing using Allied Biosystems 373 fluorescent sequencers (Perkin Elmer, Foster City, CA).

Pre Sequence Generation

The various modules used in the zymogen activation constructs are schematized in Figure 1. The bovine prolactin pre sequence signal sequence fused upstream of the FLAG epitope in a manner similar to that previously described (Ishii, et al. (1993). *J Biol Chem* 268:9780-6). This sequence module was generated by designing a series of 5 double stranded oligonucleotides having cohesive overhangs. These oligonucleotides were kinased, paired (PF-#1U with PF-#10L, PF-#2U with PF-#9L, PF-#3U with PF-#8L, PF-#4U with PF-#7L, PF-#5U with PF-#6L; Table 1), in 500 mM NaCl and annealed in 5 separate reactions. Aliquots of the annealed oligonucleotides were combined, ligated and the product subjected to PCR with primers PF-#1U and PF-#6L. This preparative reaction was performed using Amplitaq (Perkin Elmer, Foster City, CA) in the buffer supplied by the manufacturer with 10 cycles of 93 °C for 45 seconds/ 60 °C for 45 seconds/ 72 °C for 45 seconds, followed by 5 min at 72 °C. The product was digested with Eco RI and Not I and ligated into the pCDNA3 vector cleaved with Eco RI and Not I followed by dephosphorylation with calf alkaline phosphatase. An isolate, containing the desired

sequence designated prolactinFLAGpCDNA3 (PFpCDNA3) was used in subsequent manipulations. Additional pre sequences such as the human trypsinogen I and chymotrypsinogenFLAG (ChymoFLAG or CF) (Figure 1) were generated by a direct double-stranded oligonucleotide insertion using the corresponding oligonucleotides (Table 1). Since these two pre sequences are shorter than that of prolactin, the annealed duplexes were designed to contain a 5'-Eco RI and a 3'-Not I cohesive ends and thereby could be inserted into the corresponding sites of pCDNA3 directly.

Most members of the S1 protease family contain a cysteine residue just upstream from the cleavage site of the pro sequence in a conserved region. This cysteine residue (Cys-1 by chymotrypsin numbering) is disulfide bonded to another conserved cysteine within the catalytic domain (Cys-122) (Matthews, et al. (1967). *Nature (London)* 214:652-6). We will refer to this class of S1 serine proteases as type II. It is possible that the existence of this catalytic cysteine residue 122 in the disulfide-bonded state is important for specific activity and/or substrate specificity. Consequently, in order to accommodate serine proteases of this type, we synthesized the CF pre sequence that will produce recombinant proteases containing a cysteine residue just upstream of the zymogen cleavage site.

Other pre sequences are suitable for use in the present invention as pre sequences for trafficking recombinant proteins into the secretory pathway of eukaryotic cells. These often include but are not limited to translational initiation methionine residues followed by a stretch of aliphatic amino acids. Export signal sequences target newly synthesized proteins to the endoplasmic reticulum of eukaryotic cells and the plasma membrane of bacteria. Although signal sequences contain a hydrophobic core region, they show great variation in both overall length and amino acid sequence. Recently, it has become clear that this variation allows signal sequences to specify different modes of targeting and membrane insertion. In the vast majority of instances, the signal peptide does not interfere with the secreted protein function following its cleavage by the signal peptidase (Martoglio and Dobberstein (1998). *Trends Cell Biol* 8:410-415). A variety of signal sequence modules, for general use in the secretion of expressed proteins, are currently commercially available

(Invtirogen, San Diego, CA), and are suitable for use in the present invention as pre sequences.

Pro Sequence Generation

The EK cleavage site of human trypsinogen I was generated using the PCR with the two primers EK1-U and EK1-L (Table 1). The template was an EST (W40511) identified through FASTA searches (Pearson and Lipman (1988). *Proc Natl Acad Sci U. S. A.* 85:2444-8) of Db EST and obtained from the I.M.A.G.E. consortium through Genome Systems Inc., St. Louis, MO. The purified plasmid DNA of W40511 was used as a template in preparative PCR reactions, with Amplitaq (Perkin Elmer, Foster City, CA) in accordance with the manufacturer's recommendations with 15 cycles of 93 °C for 45 seconds/ 53 °C for 45 seconds/ 72 °C for 45 seconds, followed by 5 min at 72 °C. The PCR product was subcloned using the T/A vector pCR 2.1 (InVitrogen, San Diego, CA) and a clone with the desired sequence was chosen. The product was preparatively isolated by digestion using Not I and Xba I and subcloned downstream of the PF pre sequence between the Not I and Xba I sites in PFpCDNA3 to make PFEKpCDNA3. Additional pro sequences such as the FXa cleavage site and variations of the EK site (EK2 and EK3) were generated by direct double-stranded oligonucleotide insertions using the corresponding oligonucleotides. By design, these oligonucleotides once annealed would possess a 5'-Not I and a 3'-Xba I site such that they could be inserted into PFpCDNA3 or CFpCDNA3, which contain the prolactinFLAG and chymotrypsinogenFLAG pre sequences respectively, to generate a series of pre-pro sequence modules such as PFFXapCDNA3 and CFEK2pcDNA3 etc.

The other class of S1 serine proteases can be generally defined by several smaller serine proteases like trypsin, prostate specific antigen, and stratum corneum chymotryptic enzyme. This class, we will refer to as type I, lack the cysteine residue just upstream of the cleavage site yet, contain a cysteine just downstream of the zymogen activation pro sequence. In the case of these trypsin-like S1 serine proteases, this cysteine (Cys-22 by chymotrypsinogen numbering) participates in disulfide bond formation with a cysteine in the catalytic domain (Cys-157) (Stroud, et al (1974). *J Mol Biol* 83:185-208, Kossiakoff et

al. (1977). *Biochemistry* 16:654-64) and may have important consequences on catalytic activity and or substrate specificity. In order to accommodate this other type of serine protease, two more EK cleavage modules for the zymogen activation constructs were generated (Figure 2).

Thus, to analyze the activity of a particular serine protease cDNA, the appropriate combination of pre-pro sequence that corresponds to the amino acid sequence of the particular serine protease, can be used. For example, the trypsin-like type I serine proteases could be expressed from a PFEK3 pre-pro sequence while a chymotrypsin-like type II protease may be better represented by the CFEK2 pre-pro modules.

Other pro sequences, and variations of them, are suitable for use in the present invention as pro sequences for cleavage by a restriction protease for activating the inactive zymogen produced by this system. These include, but are not limited to, the cleavage sites for the restriction proteases thrombin and PreScission™ Protease (Pharmacia Biotech Inc., Piscataway, NJ).

C-terminal Affinity/Epitope Tags

Kinased, annealed double-stranded oligonucleotides, containing 5'-Xba I-and 3'-Not I cohesive ends were designed corresponding to either a stop codon, 6 histidine codons and a C-terminal stop codon (6XHISTAG), or a Hemagglutinin epitope tag with a C-terminal stop codon (HATAG) (Figure 1 and Table 1). These oligonucleotides were individually ligated between the Xba I and Not I sites in the plasmid vector pCI Neo (Promega, Madison, WI). Likewise, oligonucleotides were designed corresponding to the Hemagglutinin epitope tag but lacking a C-terminal stop codon (HA-Nonstop). This kinased annealed double-stranded oligonucleotide, containing Xba I cohesive termini, was reiteratively inserted upstream of the HATAG to generate a 3XHATAG epitope tag. In addition, the HA-Nonstop oligonucleotide was inserted upstream of the 6XHISTAG to generate a Hemagglutinin epitope/ 6XHIS affinity tag (HA6XHISTAG).

Zymogen Activation Vector Generation

The series of pre-pro sequences described above (ex. PFFXa or CFEK2 etc.) were preparatively excised from the pCDNA3 vector using Eco RI and Xba I. The FXa sequence, shown in Table 1 in particular, contains a Xba I site which becomes blocked by overlapping Dam methylation. To overcome this phenomenon, plasmid DNA of these FXa recombinants had to be transformed into and purified from a strain lacking Dam methylation (SCS110 for ex. Stratagene, La Jolla, CA) in order to cleave this site using the Xba I restriction enzyme. The pre-pro sequences were ligated into the various C-terminal epitope or affinity tagged pCIneo constructs between their 5'-Eco RI and 3'-Xba I sites. Thus, these constructs all feature a pre sequence (prolactin FLAG, PF; chymotrypsinogenFLAG, CF; or trypsinogen, T) to direct secretion in-frame with a pro sequence recognized by a restriction protease EK (sites EK1 EK2 EK3); or factor Xa (site FXa), to permit the post-translational cleavage for zymogen activation. A unique Xba I restriction enzyme site immediately upstream of the epitope/affinity tags, described above, separates these pre-pro combinations (Figure 2). Due to the nature of the design, the Xba I site is critical to these vectors, and was chosen based on several criteria as follows. These include the observation that the "6-cutter" (a restriction enzyme recognizing 6 nucleotide bases in its specific cleavage site) restriction enzyme Xba I site is found infrequently within cDNAs which greatly minimizes labor-intensive cloning steps in the generation of cDNA expression constructs for general use. Additionally, should one or more Xba I sites exist within a particular cDNA sequence one desires to insert into this vector, two other restriction enzymes (Spe I and Nhe I) are also rare 6-cutters which give rise to Xba I compatible cohesive ends. It should be noted that in this series of zymogen activation constructs, the translational register of the pre-pro sequences is distinct from that of the epitope/affinity tags. The resulting recombinants comprise a series of mammalian zymogen activation constructs in the pCIneo background. For increased levels of expression, these pre-pro-epitope modules were individually shuttled into vectors capable of expression in Drosophila S2 cells. This was accomplished by preparatively isolating the individual pre-pro-Xba I-epitope/affinity-tag modules by digesting the mammalian pCI Neo zymogen activation constructs with 5'-Eco RI and 3'-Hinc II. These modules were then inserted into the Eco RI

and Hinc II sites of either an inducible Drosophila vector pRM63 containing the metallothionein promoter, or the constitutive Drosophila vector pFLEX64 containing the actin 5c promoter.

EXAMPLE 2

Acquisition of Serine Protease cDNAs

Acquisition of a full length cDNA corresponding to the serine protease prostasin

The full length cDNA for prostasin (Yu, et al. (1995). J Biol Chem 270:13483-9) was identified through FASTA searches of Db EST (Genbank accession number AA205604) and obtained from the I.M.A.G.E. consortium through Genome Systems, Inc., St. Louis, MO. The clone was sequenced for confirmation.

Acquisition of a full length cDNA corresponding to the novel protease O

A putative full-length clone of a novel serine protease (Yoshida, et al., (1998). Biochim. Biophys. Acta, 1399:225-228), designated protease O, was cloned and sequenced for confirmation.

Acquisition of a full length cDNA corresponding to the human orthologue of protease neuropsin

A partial clone with homology to the murine neuropsin (Chen, et al. (1995). J Neurosci 15:5088-97) was also identified (Yoshida, et al., (1998). Gene, 213:9-16). The full-length cDNA of human neuropsin was obtained by screening a Uni-ZAP keratinocyte library, followed by *in vivo* excision and sequence analysis of positive purified plaques.

Acquisition of a full length cDNA corresponding to protease F/ESP-1

Homology searches identified a novel serine protease, we designated proteases F, within sequence nucleotide databases. An EST containing the full length cDNA for protease F was identified through FASTA searches of Db EST (Genbank accession

number AA159101) and obtained from the I.M.A.G.E. consortium through Genome Systems, Inc., St. Louis, MO. The clone was sequenced for confirmation. The nucleotide and deduced amino acid sequences were subsequently published (Inoue, et al. (1998). Biochem. Biophys. Res. Commun. 252:307-312) during the proceeding of our investigations.

Acquisition of the protease MH2/Prostase catalytic domain

Homology searches identified a novel serine protease we designated proteases MH2 within sequence nucleotide databases. This particular serine protease was of interest since expression profiling had indicated prostate specific expression. We employed the 3' and 5' rapid amplification of cDNA ends (RACE) method in an attempt the isolate the full length protease MH2 cDNA using prostate marathon ready cDNA and random primed 5'-adapter-linked prostate cDNA (Clontech, Palo Alto, CA). Despite numerous attempts, we were only able to obtain clones which contained the protease MH2 catalytic domain and lacked the initiation methionine and signal sequence. The nucleotide and deduced amino acid sequences were subsequently published (Nelson et al. (1999). Proc. Natl. Acad. Sci. U. S. A. 96:3114-3119) during the proceeding of our investigations.

General plasmid manipulation

The purified plasmid DNA of these serine protease cDNAs was used as a template in 100 ul preparative PCR reactions with Amplitaq (Perkin Elmer, Foster City, CA) or Pfu DNA polymerase (Stratagene, La Jolla, CA) in accordance with the manufacturer's recommendations. Typically, reactions were run at 18 cycles of 93 °C for 30 seconds/ 53 to 65 °C for 30 seconds/ 72 °C for 90 seconds, followed by 5 min at 72 °C using the *Pfu* DNA polymerase. The annealing temperatures used were determined for the particular construct by the PrimerSelect 3.11 program (DNASTAR Inc., Madison, WI). The primers of the respective serine proteases (Table 1), containing Xba I cleavable ends, were designed to flank the catalytic domains of these

three proteases and generate Xba I catalytic cassettes (Figure 1). Since the protease prostasin is initially thought to be C-terminally membrane bound, and subsequently rendered soluble through proteolysis following secretion (Yu, et al. (1995). J Biol Chem 270:13483-9), a soluble form of prostasin was generated. This was accomplished by excluding the C-terminal 29 amino acids in the prostasin catalytic cassette by designing the C-terminal Xba I primer (prostasin(SOL) Xba-L, Table 1) to a position immediately upstream from the hydrophobic stretch of amino acids thought to represent a membrane tether.

The preparative PCR products were phenol/CHCl3 (1:1) extracted once, CHCl3 extracted, and then EtOH precipitated with glycogen (Boehringer-Mannheim Corp., Indianapolis, IN) carrier. The precipitated pellets were rinsed with 70 % EtOH, dried by vacuum, and resuspended in 80 ul H20, 10 ul 10 restriction buffer number 2 and 1 ul 100x BSA (New England Biolabs, Beverly, MA). The products were digested for at least 3 hours at 37 oC with 200 units Xba I restriction enzyme (New England Biolabs, Beverly, MA). The Xba I digested products were phenol/CHCl3 (1:1) extracted once, CHCl3 extracted, EtOH precipitated rinsed with 70 % EtOH, and dried by vacuum. For purification from contaminating template plasmid DNA, the products were electrophoresed through 1.0 % low melting temperature agarose (Life Technologies, Gaithersberg, MD) gels in TAE buffer (40 mM Tris-Acetate, 1 mM EDTA pH 8.3) and excised from the gel. Aliquots of the excised products were routinely used for in-gel ligations with the appropriate Xba I digested, dephosphorylated and gel purified, zymogen activation vector. These cassettes once inserted, in the correct orientation, placed them in the proper translational register with the NH2-terminal prepro sequence and C-terminal/epitope affinity tag. PCR products directly cloned, as described above, were sequenced for confirmation. Only clones having confirmed sequences were chosen to isolate the Xba I catalytic cassette for subsequent subcloning into additional vectors of the series when desired.

EXAMPLE 3

Expression of Recombinant Serine Proteases in Drosophila S2 Cells

The recombinant bacmid containing the zymogen activated constructs were prepared from bacterial transformation, selection, growth, purification and PCR confirmation in accordance with the manufacturer's recommendations. Cultured Sf9 insect cells (ATCC CRL-1711) were transfected with purified bacmid DNA and several days later, conditioned media containing recombinant zymogen activated baculovirus was collected for viral stock amplification. Sf9 cells growing in Sf-900 II SFM at a density of $2X10^6$/ml were infected at a multiplicity of infection of 2 at 27 °C for 80 hours, and cell pellets were collected for purification of the zymogen activated constructs.

EXAMPLE 4

Purification, and Activation of Recombinant Serine Proteases

Cells were lysed on ice in 20 mM Tris (pH7.4), 150 mM NaCl, 1% Triton X-100, 1 mM EDTA, 1 mM EGTA, 1 mM PMSF, leupeptin (1 μg/ml), and pepstatin (1 μg/ml). Cell lysates were mixed with anti-FLAG M2 affinity gel (Eastman Kodak Co., New Haven, CT) and bound at 4 °C for 3 hours with gentle rotation. The zymogen-bound resin was washed 3 times with TBS buffer (50 mM Tris-HCl, 150 mM NaCl at a final pH of 7.5), and eluted by competition with FLAG peptide (100 μg/ml) in TBS buffer. The eluted zymogen was dialyzed overnight against TBS in Spectra/Por membrane (MWCO: 12,000-14,000) (Spectra Medical Industries, Inc., Huston, TX). Ni-NTA (150 μl of a 50 % slurry/per 100 μg of zymogen) (Qiagen, Valencia, CA) was added to 5 ml the dialyzed sample and mixed by shaking at 4 °C for 60 minutes The zymogen-bound resin was washed 3 times with wash buffer [10 mM Tris-HCl (pH 8.0), 300 mM NaCl, and 15 mM imidazole], followed by with a 1.5 ml wash with ds H_2O. Zymogen cleavage was carried out by adding enterokinase (10 U per 50 μg of zymogen) (Novagen, Inc., Madison WI; or Sigma, St. Louis, MO) to the zymogen-bound Ni-NTA beads in a small volume at room temperature overnight

with gentle shaking in a buffer containing 20 mM Tris-HCl (pH 7.4), 50 mM NaCl, and 2.0 mM $CaCl_2$. The resin was then washed twice with 1.5 ml wash buffer. The activated protease was eluted with elution buffer [20 mM Tris-HCl (pH 7.8), 250 mM NaCl, and 250 mM imidazole]. Eluted protein concentration was determined by a Micro BCA Kit (Pierce, Rockford, IL) using bovine serum albumin as a standard. Amidolytic activities of the activated protease was monitored by release of para-nitroaniline (pNA) from the synthetic substrates indicated in Table 2. The chromogenic substrates used in these studies were all commercially available (Bachem California Inc., Torrance, PA; American Diagnostica Inc., Greenwich, CT; Kabi Pharmacia Hepar Inc., Franklin, OH). Assay mixtures contained chromogenic substrates at 500 uM and 10 mM Tris-HCl (pH 7.8), 25 mM NaCl, and 25 mM imidazole. Release of pNA was measured over 120 minutes at 37 °C on a micro-plate reader (Molecular Devices, Menlo Park, CA) with a 405 nm absorbance filter. The initial reaction rates (Vmax, mOD/min) were determined from plots of absorbance versus time using Softmax (Molecular Devices, Menlo Park, CA). The specific activities (nmole pNA produced /min/ug protein) of the activated proteases for the various substrates are presented in Table 2. No measurable chromogenic amidolytic activity was detected with the purified unactivated zymogens.

EXAMPLE 5

Electrophoresis and Western Blotting Detection of Recombinant Serine Proteases

Samples of the purified zymogens or activated proteases, denatured in the presence or absence of the reducing agent dithiothreitol (DTT), were analyzed by SDS-PAGE (Bio Rad, Hercules CA) stained with Coomassie Brilliant Blue. For Western Blotting, the Flag-tagged serine proteases expressed from transient or stable S2 cells were detected with anti-Flag M2 antibody (Babco, Richmond, CA). The secondary antibody was a goat-anti-mouse IgG (H+L), horseradish peroxidase-linked F(ab')2 fragment, (Boehringer Mannheim Corp., Indianapolis, IN) and was detected by the ECL kit (Amersham, Arlington Heights, IL). Figure 7 demonstrates PFEK2-prostasin-6XHIS function by demonstrating the quantitative

cleavage of the expressed and purified zymogen to generate the processed and activated protease. Since the FLAG epitope is located just upstream of the of the EK pro sequence, cleavage with EK generates a FLAG-containing polypeptide which is too small to be retained in the polyacrylamide gel, and is therefore not detected in the +EK lanes. Also shown in panel B, the untreated or EK digested PFEK2-prostasin-6XHIS was denatured in the absence of DTT, in order to retain disulfide bonds, prior to electrophoresis (lanes 3 and 4). Although equivalent amounts of sample were loaded into each lane of the gel in the Western blot of B, the anti-FLAG MoAb M2 appears to detect proteins better when pretreated with DTT (compare lane B1 with B3). Figure 8 demonstrates CFEK2-prostasin-6XHIS function by demonstrating the quantitative cleavage of the expressed and purified zymogen to generate the processed and activated protease. Since the FLAG epitope is located just upstream of the of the EK2 pro sequence, cleavage with EK generates a FLAG-containing polypeptide which is too small to be retained in the polyacrylamide gel, and is therefore not detected in the +EK lanes. Also shown in panel B, the untreated or EK digested CFEK2-prostasin-6XHIS was denatured in the absence of DTT, in order to retain disulfide bonds, prior to electrophoresis (lanes 3 and 4). Of significance in lane 4 is the retention of the FLAG epitope indicating the formation of a disulfide bond between the cysteine in the CF pre sequence with a cysteine in the catalytic domain of prostasin which is presumably Cys-122 (chymotrypsin numbering). Retention of the FLAG epitope, following EK cleavage and denaturation without DTT, is not observed using the prolactin pre sequence which lacks a cysteine residue (Compare lane 4 of Figure 7 with lane 4 of Figure 8). This documents that the CF pre sequence is capable of forming a light chain, that is disulfide bonded to the heavy catalytic chain of the recombinant serine proteases, when expressed in this system. It appears that in the absence of the reducing agent DTT, the EK cleaved polypeptides have a reproducibly decreased mobility in the gel (compare lane B3 with B4). Figure 9 demonstrates function of PFEK1-neuropsin-6XHIS by demonstrating quantitative cleavage of the expressed and purified zymogen to generate the processed and activated protease. Figure 10 demonstrates function of PFEK1-protease O-6XHIS by demonstrating quantitative cleavage of the expressed and purified zymogen to generate the

processed and activated protease. Figure 11 demonstrates function of PFEK1-protease F-6XHIS by demonstrating quantitative cleavage of the expressed and purified zymogen to generate the processed and activated protease. Figure 12 demonstrates function of PFEK1-protease MH2-6XHIS by demonstrating quantitative cleavage of the expressed and purified zymogen to generate the processed and activated protease.

EXAMPLE 6

Chromogenic Assay

Amidolytic activities of the activated serine proteases are monitored by release of para-nitroaniline (pNA) from synthetic substrates that are commercially available (Bachem California Inc., Torrance, PA; American Diagnostica Inc., Greenwich, CT; Kabi Pharmacia Hepar Inc., Franklin, OH). Assay mixtures contain chromogenic substrates in 500 uM and 10 mM TRIS-HCl (pH 7.8), 25 mM NaCl, and 25 mM imidazole. Release of pNA is measured over 120 min at 37 °C on a micro-plate reader (Molecular Devices, Menlo Park, CA) with a 405 nm absorbance filter. The initial reaction rates (Vmax, mOD/min) are determined from plots of absorbance versus time using Softmax (Molecular Devices, Menlo Park, CA). Compounds that modulate a serine protease of the present invention are identified through screening for the acceleration, or more commonly, the inhibition of the proteolytic activity. Although in the present case chromogenic activity is monitored by an increase in absorbance, fluorogenic assays or other methods such as FRET to measure proteolytic activity as mentioned above, can be employed. Compounds are dissolved in an appropriate solvent, such as DMF, DMSO, methanol, and diluted in water to a range of concentrations usually not exceeding 100 uM and are typically tested, though not limited to, a concentration of 1000-fold the concentration of protease. The compounds are then mixed with the protein stock solution, prior to addition to the reaction mixture. Alternatively, the protein and compound solutions may be added independently to the reaction mixture, with the compound being added either prior to, or immediately after, the addition of the protease protein.

Table 1

SEQ.ID ,NO.:	Oligo Name	Sequence
15	Stop-U	CTAGATAGC
16	Stop-L	GGCCGCTAT
17	HA-Stop-U	CTAGATACCCCTACGATGTGCCCGATTACGCCTAGC
18	HA-Stop-L	GGCCGCTAGGCGTAATCGGGCACATCGTAGGGGTAT
19	HA-Nonstop-U	CTAGATACCCCTACGATGTGCCCGATTACGCCG
20	HA-Nonstop-L	CTAGCGGCGTAATCGGGCACATCGTAGGGGTAT
21	6XHIS-U	CTAGACATCACCATCACCATCACTAGC
22	6XHIS-L	GGCCGCTAGTGATGGTGATGGTGATGT
23	PF-#1U	TGAATTCACCACCATGGACAGCAAAGGTTCGTCG
24	PF-#2U	CAGAAAGGGTCCCGCCTGCTCCTGCTGCTG
25	PF-#3U	GTGGTGTCAAATCTACTCTTGTGCCAGGGT
26	PF-#4U	GTGGTCTCCGACTACAAGGACGACGACGAC
27	PF-#5U	GTGGACGCGGCCGCATTATTA
28	PF-#6L	TAATAATGCGGCCGCGTCCACGTCGTCGTCGTCCT
29	PF-#7L	TGTAGTCGGAGACCACACCCT
30	PF-#8L	GGCACAAGAGTAGATTTGACACCACCAGCA
31	PF-#9L	GCAGGAGCAGGCGGGACCCTTTCTGCGACG
32	PF-#10L	AACCTTTGCTGTCCATGGTGGTGAATTCA
33	TrypIPre-U	AATTCACCATGAATCCACTCCTGATCCTTACCTTTGTGGC
34	TrypIPre-L	GGCCGCCACAAAGGTAAGGATCAGGAGTGGATTCATGGTG
35	CF-#1U	AATTCACCACCATGGCTTTCCTCTGGCTCCTCTCCTGCTGGG CCCTCCTGGGTAC
36	CF-#2L	CCAGGAGGGCCCAGCAGGAGAGGAGCCAGAGGAAAGCCATGG TGGTG
37	CF-#3U	CACCTTCGGCTGCGGGGTCCCCGACTACAAGGACGACGACGA CGC
38	CF-#4L	GGCCGCGTCGTCGTCGTCCTTGTAGTCGGGGACCCCGCAGCC GAAGGTGGTAC

39	EK1-U	GTGGCGGCCGCTCTTGCTGCCCCCTTTGA
40	EK1-L	TTCTCTAGACAGTTGTAGCCCCCAACGA
41	EK2-U	GGCCGCTCTTGCTGCCCCCTTTGATGATGATGACAAGATCGT TGGGGGCTATGCT
42	EK2-L	CTAGAGCATAGCCCCCAACGATCTTGTCATCATCATCAAAGG GGGCAGCAAGAGC
43	EK3-U	GGCCGCTCTTGCTGCCCCCTTTGATGATGATGACAAGATCGT TGGGGGCTATTGT
44	EK3-L	CTAGACAATAGCCCCCAACGATCTTGTCATCATCATCAAAGG GGGCAGCAAGAGC
45	FXa-U	GGCCGCTCTTGCTGCCCCCTTTATCGAGGGGCGCATTGTGGA GGGCTCGGAT
46	FXa-L	CTAGATCCGAGCCCTCCACAATGCGCCCCTCGATAAAGGGGG CAGCAAGAGC
47	prostasin Xba-U	AGCAGTCTAGAGGCCGGTCAGTGGCCCTGGCA
48	prostasin(SOL) Xba-L	GCTGGTCTAGAGCTGAAGGCCAGGTGGC
49	neuropsin Xba-U	GGTATCTAGAGCCCTTGCTGCCTATGATC
50	neuropsin Xba-L	ACTGTCTAGAACCCCATTCGCAGCCTTGGC
51	protease O Xba-U	TCGATCTAGAAAAGCACTCCCAGCCCTGGCAG
52	protease O Xba-L	GTCCTCTAGAATTGTTCTTCATCGTCTCCTGG

Protease cDNA	Genbank Acc.#	
h Trypsinogen I	W40511	
h Prostasin	AA205604	
h Neuropsin		2604309
h Protease O		2723646

Table 2

Recombinant Protease	H-D-Pro-HHT-Arg-pNA	H-D-Lys(CBO)-Pro-Arg-pNA	H-D-Val-Leu-Lys-pNA	H-DL-Val-Leu-Arg-pNA
PFEK2-prostasin-6XHIS	0.055±0.002	0.870±0.022	N.D.	0.251±0.005
CFEK2-prostasin-6XHIS	0.116±0.011	1.317±0.024	N.D.	0.384±0.003
PFEK1-neuropsin-6XHIS	0.463±0.014	0.731±0.004	0.158±0.001	0.938±0.002
PFEK1-protease O-6XHIS	0.058±0.002	0.022±0.000	N.D.	0.006±0.000
PFEK-MH2-6XHIS	0.052±0.000	0.893±0.067	0.121±0.054	0.058±0.002
CFEK2-Prot.F-6XHIS	0.016±0.001	0.045±0.006	N.D.	N.D.

References Cited

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403-10.

Chen, Z.-L., Yoshida, S., Kato, K., Momota, Y., Suzuki, J., Tanaka, T., Ito, J., Nishino, H., Aimoto, S., Kiyama, H., and Shiosaka, S. (1995). Expression and activity-dependent changes of a novel limbic-serine protease gene in the hippocampus. J. Neurosci. *15*, 5088-97.

Davie, E. W., Fujikawa, K., and Kisiel, W. (1991). The coagulation cascade: initiation, maintenance, and regulation. Biochemistry *30*, 10363-70.

Hansson, L., Stroemqvist, M., Baeckman, A., Wallbrandt, P., Carlstein, A., and Egelrud, T. (1994). Cloning, expression, and characterization of stratum corneum chymotryptic enzyme. A skin-specific human serine proteinase. J. Biol. Chem. *269*, 19420-6.

Huber, R., and Bode, W. (1978). Structural basis of the activation and action of trypsin. Acc. Chem. Res. *11*, 114-22.

Inoue, M., Kanbe, N., Kurosawa, M., and Kido, H. (1998). Cloning and tissue distribution of a novel serine protease esp-1 from human eosinophils. Biochem. Biophys. Res. Commun. 252, 307-312.

Ishii, K., Hein, L., Kobilka, B., and Coughlin, S. R. (1993). Kinetics of thrombin receptor cleavage on intact cells. Relation to signaling. J. Biol. Chem. *268*, 9780-6.

Kitamoto, Y., Yuan, X., Wu, Q., McCourt, D. W., and Sadler, J. E. (1994). Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of a distinctive assortment of domains. Proc. Natl. Acad. Sci. U. S. A. *91*, 7588-92.

Kossiakoff, A. A., Chambers, J. L., Kay, L. M., and Stroud, R. M. (1977). Structure of bovine trypsinogen at 1.9 .ANG. resolution. Biochemistry *16*, 654-64.

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. *157*, 105-32.

Leytus, S. P., Loeb, K. R., Hagen, F. S., Kurachi, K., and Davie, E. W. (1988). A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells. Biochemistry *27*, 1067-74.

Little, S. P., Dixon, E. P., Norris, F., Buckley, W., Becker, G. W., Johnson, M., Dobbins, J. R., Wyrick, T., Miller, J. R., Mackellar, W., Hepburn, D., Corvalan, J., Mcclure, D., Liu, X., Stephenson, D., Clemens, J., and Johnstone, E. M. (1997). Zyme, a novel and potentially amyloidogenic enzyme cDNA isolated from Alzheimer's disease brain. J. Biol. Chem. *272*, 25135-25142.

Martoglio, B., and Dobberstein, B. (1998). Signal sequences: more than just greasy peptides. Trends Cell Biol. *8*, 410-415.

Matthews, B. W., Sigler, P. B., Henderson, R., and Blow, D. M. (1967). Three-dimensional structure of tosyl-.alpha.-chymotrypsin. Nature (London) *214*, 652-6.

Nelson, P. S., Gan, L., Ferguson, C., Moss, P., Gelinas, R., Hood, L., and Wang, K. (1999). Molecular cloning and characterization of prostase, an androgen-regulated serine protease with prostate-restricted expression. Proc. Natl. Acad. Sci. U. S. A. 96, 3114-3119.

Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. U. S. A. *85*, 2444-8.

Proud, D., and Kaplan, A. P. (1988). Kinin formation: mechanisms and role in inflammatory disorders. Annu. Rev. Immunol. *6,*, 49-83.

Rawlings, N. D., and Barrett, A. J. (1994). Families of serine peptidases. Methods Enzymol. *244*, 19-61.

Reid, K. B. M., and Porter, R. R. (1981). The proteolytic activation systems of complement. Annual Review of Biochemistry *50*, 433-464.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed.: Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Stroud, R. M., Kay, L. M., and Dickerson, R. E. (1974). Structure of bovine trypsin. Electron density maps of the inhibited enzyme at 5 .ang. and 2.7 .ang. resolution. J. Mol. Biol. *83*, 185-208.

Tachias, K., and Madison, E. L. (1996). Converting tissue-type plasminogen activator into a zymogen. J. Biol. Chem. *271*, 28749-28752.

Takayama, T. K., Fujikawa, K., and Davie, E. W. (1997). Characterization of the precursor of prostate-specific antigen Activation by trypsin and by human glandular kallikrein. J. Biol. Chem. *272*, 21582-21588.

Wang, Z.-m., Rubin, H., and Schechter, N. M. (1995). Production of active recombinant human chymase from a construct containing the enterokinase cleavage site of trypsinogen in place of the native propeptide sequence. Biol. Chem. Hoppe-Seyler *376*, 681-4.

Wigler, M., Silverstein, S., Lee, L.-S., Pellicer, A., Cheng, Y.-C., and Axel, R. (1977). Transfer of purified Herpes virus thymidine kinase gene to cultured mouse cells. Cell (Cambridge, Mass.) *11*, 223-32.

Yamashiro, K., Tsuruoka, N., Kodama, S., Tsujimoto, M., Yamamura, Y., Tanaka, T., Nakazato, H., and Yamaguchi, N. (1997). Molecular cloning of a novel trypsin-like serine protease (neurosin) preferentially expressed in brain. Biochim. Biophys. Acta *1350*, 11-14.

Yoshida, S., Taniguchi, M., Hirata, A., and Shiosaka, S. (1998). Sequence analysis and expression of human neuropsin cDNA and gene. Gene *213*, 9-16.

Yoshida, S., Taniguchi, M., Suemoto, T., Oka, T., He, X., and Shiosaka, S. (1998). cDNA cloning and expression of a novel serine protease, TLSP1. Biochim. Biophys. Acta *1399*, 225-228.

Yu, J. X., Chao, L., and Chao, J. (1994). Prostasin is a novel human serine proteinase from seminal fluid. Purification, tissue distribution, and localization in prostate gland. J. Biol. Chem. 269, 18843-8.

Yu, J. X., Chao, L., and Chao, J. (1995). Molecular cloning, tissue-specific expression, and cellular localization of human prostasin mRNA. J. Biol. Chem. *270*, 13483-9.

WHAT IS CLAIMED IS:

1. An expression vector comprising, in frame and in order, a pre sequence, a pro sequence, and a cloning site for in frame insertion of a catalytic domain cassette.

2. The expression vector of claim 1, additionally comprising a tag sequence in frame with the cloning site.

3. The expression vector of claim 2 wherein said vector comprises a DNA sequence selected from the group consisting of SEQ.ID.NO.:1, SEQ.ID.NO.:2, SEQ.ID.NO.:3, SEQ.ID.NO.:4, SEQ.ID.NO.:5, and SEQ.ID.NO.:6.

4. The expression vector of claim 1, wherein said vector contains a catalytic domain cassette inserted in frame into the cloning site.

5. A recombinant host cell containing the expression vector of claim 4.

6. A process for expression of a zymogen, comprising:
(a) transferring the expression vector of claim 4 into suitable host cells; and
(b) culturing the host cells of step (a) under conditions that allow expression of the zymogen expression vector.

7. The process of claim 6, wherein said expression vector comprises a nucleotide sequence selected from a group consisting of SEQ.ID.NO.:1, SEQ.ID.NO.:2, SEQ.ID.NO.:3, SEQ.ID.NO.:4, SEQ.ID.NO.:5, SEQ.ID.NO.:6, SEQ.ID.NO.:7, SEQ.ID.NO.:8, SEQ.ID.NO.:9, SEQ.ID.NO.:10, SEQ.ID.NO.:59, and SEQ.ID.NO.:60.

8. A serine protease catalytic domain produced from a recombinant host cell containing the expression vector of claim 4, which functions as a serine protease when said protein is cleaved at the pre sequence.

9. A serine protease catalytic domain produced from a recombinant host cell containing the expression vector of claim 8 wherein the amino acid sequence is selected from a group consisting of SEQ.ID.NO.:11, SEQ.ID.NO.:12, SEQ.ID.NO.:13, SEQ.ID.NO.:14, SEQ.ID.NO.:53, SEQ.ID.NO.:54, and functional derivatives thereof.

10. The protease of claim 8, wherein said protease is bound to Ni-NTA silica or Ni-NTA agarose beads.

11. A method for identifying compounds that modulate the activity of a protease expressed from the expression vector of claim 4, comprising:

(a) combining a modulator of protease activity, protease protein, and a labeled substrate; and

(b) measuring a change in the labeled substrate.

12. The method of claim 11 wherein the labeled substrate is selected from the group consisting of flourogenic, colormetric, radiometric, and fluorescent resonance energy transfer (FRET).

13. A compound active in the method of Claim 11, wherein said compound is a modulator of a serine protease catalytic domain.

14. A compound active in the method of Claim 11, wherein the effect of the modulator on the protease is inhibiting or enhancing its enzymatic activity.

15. A compound active in the method of Claim 11, wherein the effect of the modulator on the protease is stimulation or inhibition of proteolysis mediated by the expressed catalytic domain.

16. A pharmaceutical composition comprising a compound of Claim 13.

17. A pharmaceutical composition comprising a compound of Claim 13, wherein said compound is a modulator of a protease selected from the group consisting of SEQ.ID.NO.11, SEQ.ID.NO.12, SEQ.ID.NO.13, SEQ.ID.NO.14, SEQ.ID.NO.53, SEQ.ID.NO.54, and functional derivatives thereof.

18. A method of treating a patient in need of such treatment for a condition that is mediated by a protease, comprising administration of the compound of Claim 13.

19. A kit comprising the expression vector selected from a group consisting of the expression vector of claim 1, the expression vector of claim 4, and functional derivatives thereof.

20. A kit comprising the nucleic acid sequence selected from the group consisting of, SEQ.ID.NO.:1, SEQ.ID.NO.:2, SEQ.ID.NO.:3, SEQ.ID.NO.:4, SEQ.ID.NO.:5, SEQ.ID.NO.:6, SEQ.ID.NO.:7, SEQ.ID.NO.:8, SEQ.ID.NO.:9, SEQ.ID.NO.:10, SEQ.ID.NO.:59, SEQ.ID.NO.:60 and fragments thereof.

21. A kit comprising a serine protease protein selected from the group consisting of, SEQ.ID.NO.:11, SEQ.ID.NO.:12, SEQ.ID.NO.:13, SEQ.ID.NO.:14, SEQ.ID.NO.:53, and SEQ.ID.NO.:54.

22 A pharmaceutical composition comprising the serine protease catalytic domain of claim 9.

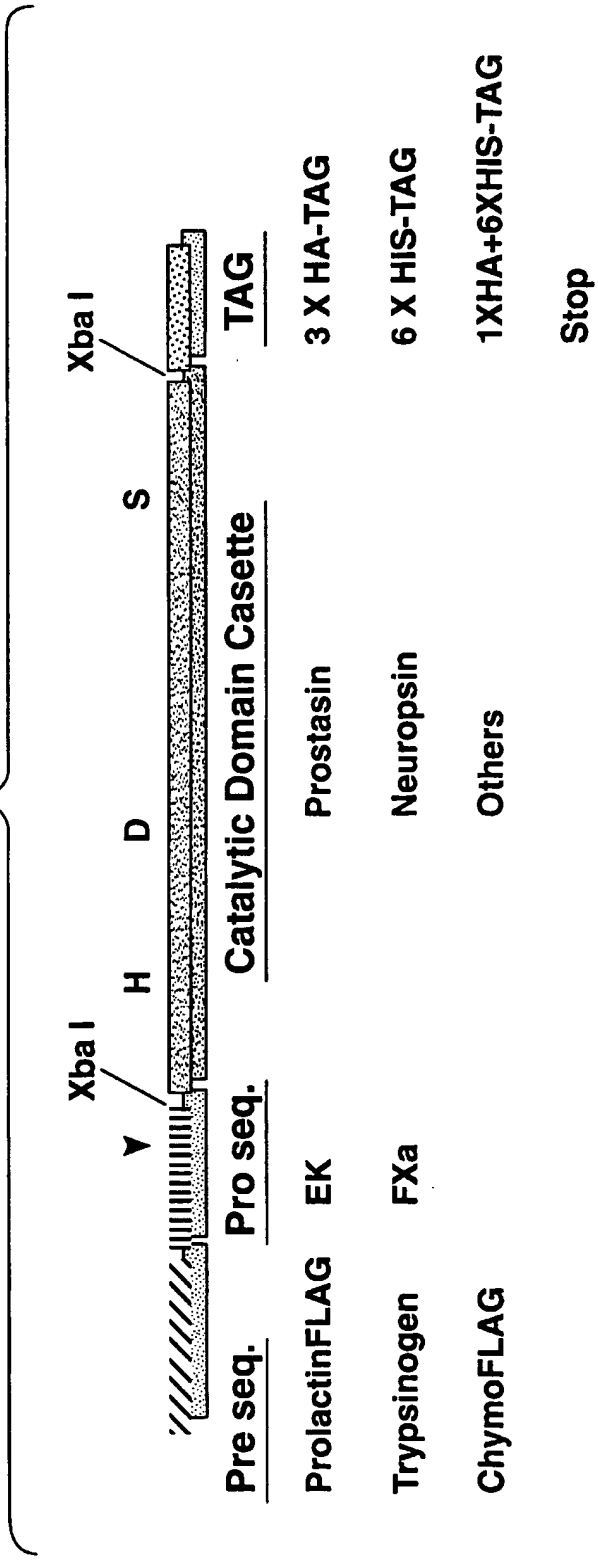
23. The pharmaceutical composition of claim 24 wherein said composition is a topical skin care composition.

24. A non-pharmaceutical composition comprising the serine protease catalytic domain of claim 9.

25. The non-pharmaceutical composition of claim 23 wherein the composition is selected from the group consisting of a laundry detergent, shampoo, hard surface cleaning compositions, and dish-care cleaning compositions.

26. A method of treating, either prophylactically or acutely, an imbalance of desquamation comprising topical application of the composition of claim 23.

FIG. 1


Xba I
H
D
S
Xba I
Pre seq.
Pro seq.
Catalytic Domain Casette
TAG
ProlactinFLAG
EK
Prostasin
3 X HA-TAG
Trypsinogen
FXa
Neuropsin
6 X HIS-TAG
ChymoFLAG
Others
1XHA+6XHIS-TAG
Stop

SEQ.ID.NO.:1

FIG. 2(A)

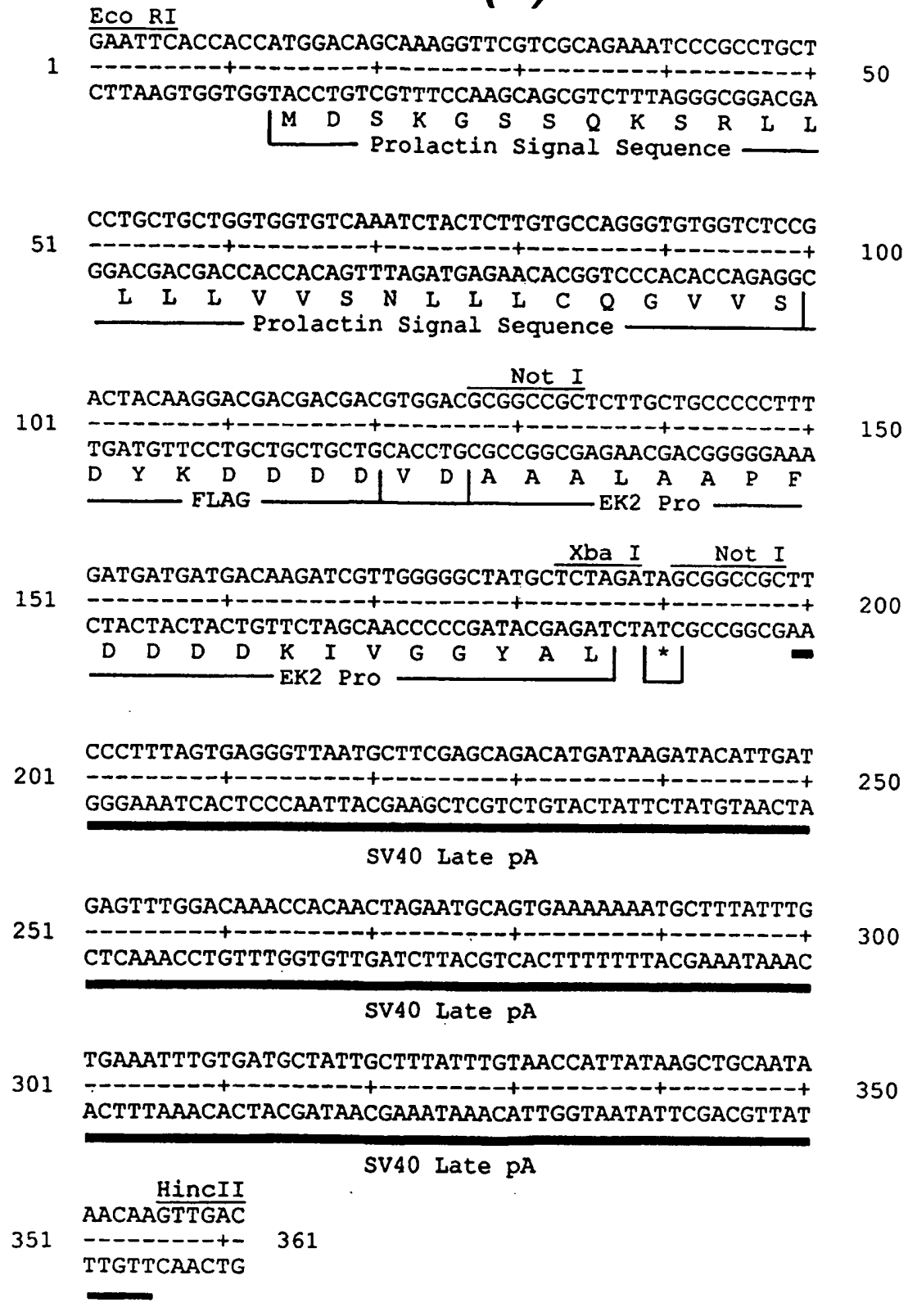


FIG. 2(B)

SEQ.ID.NO.:2

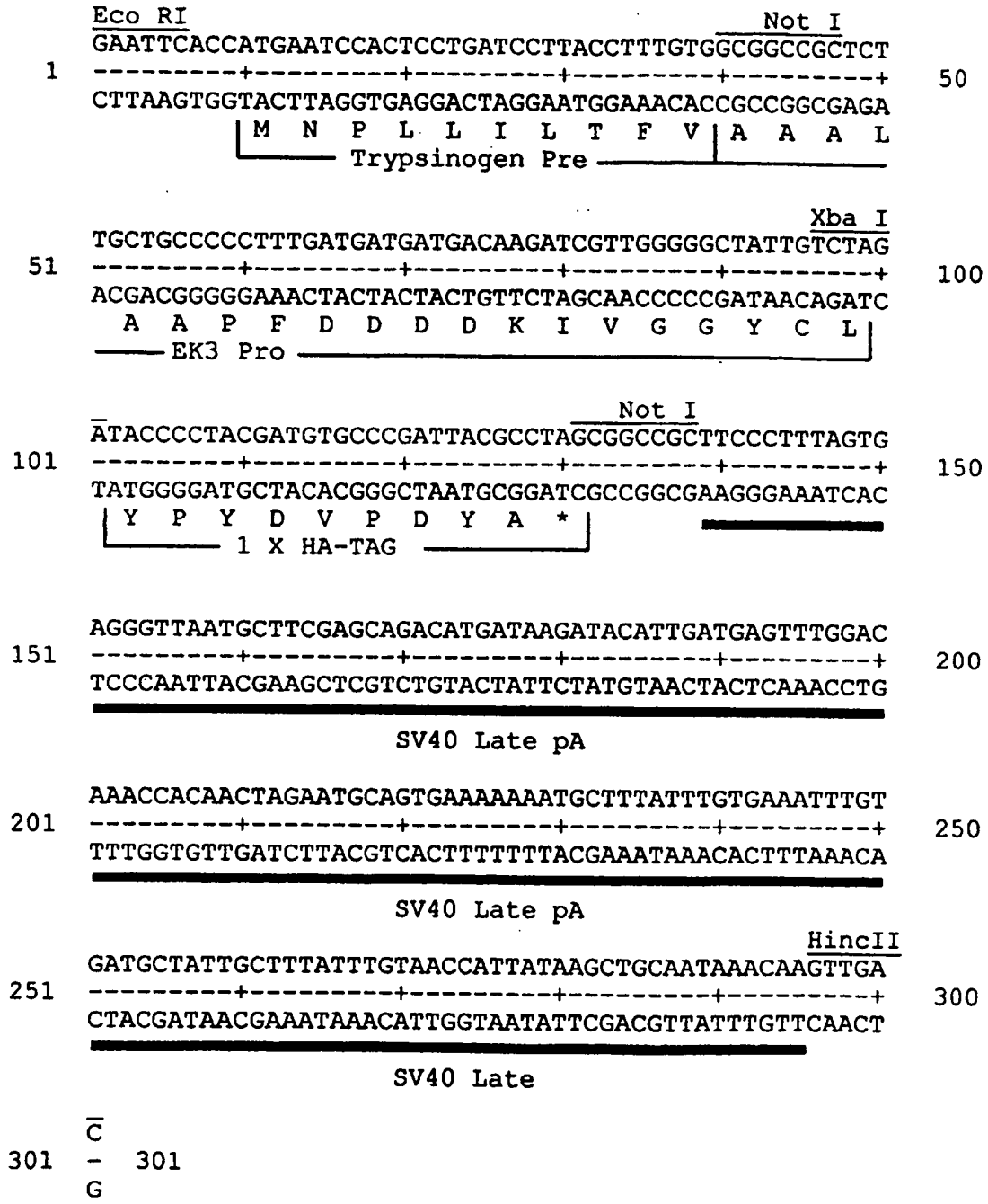


FIG. 3(D)

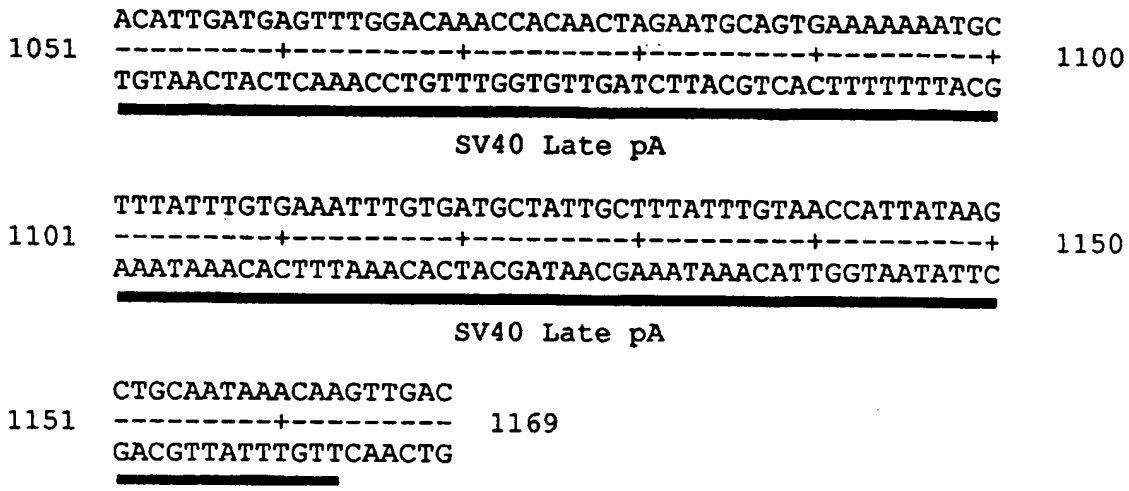

ACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGC
1051 ---------+---------+---------+---------+---------+ 1100
TGTAACTACTCAAACCTGTTTGGTGTTGATCTTACGTCACTTTTTTTACG
SV40 Late pA
TTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAG
1101 ---------+---------+---------+---------+---------+ 1150
AAATAAACACTTTAAACACTACGATAACGAAATAAACATTGGTAATATTC
SV40 Late pA
CTGCAATAAACAAGTTGAC
1151 ---------+--------- 1169
GACGTTATTTGTTCAACTG

FIG. 2(C)

SEQ.ID.NO.:3

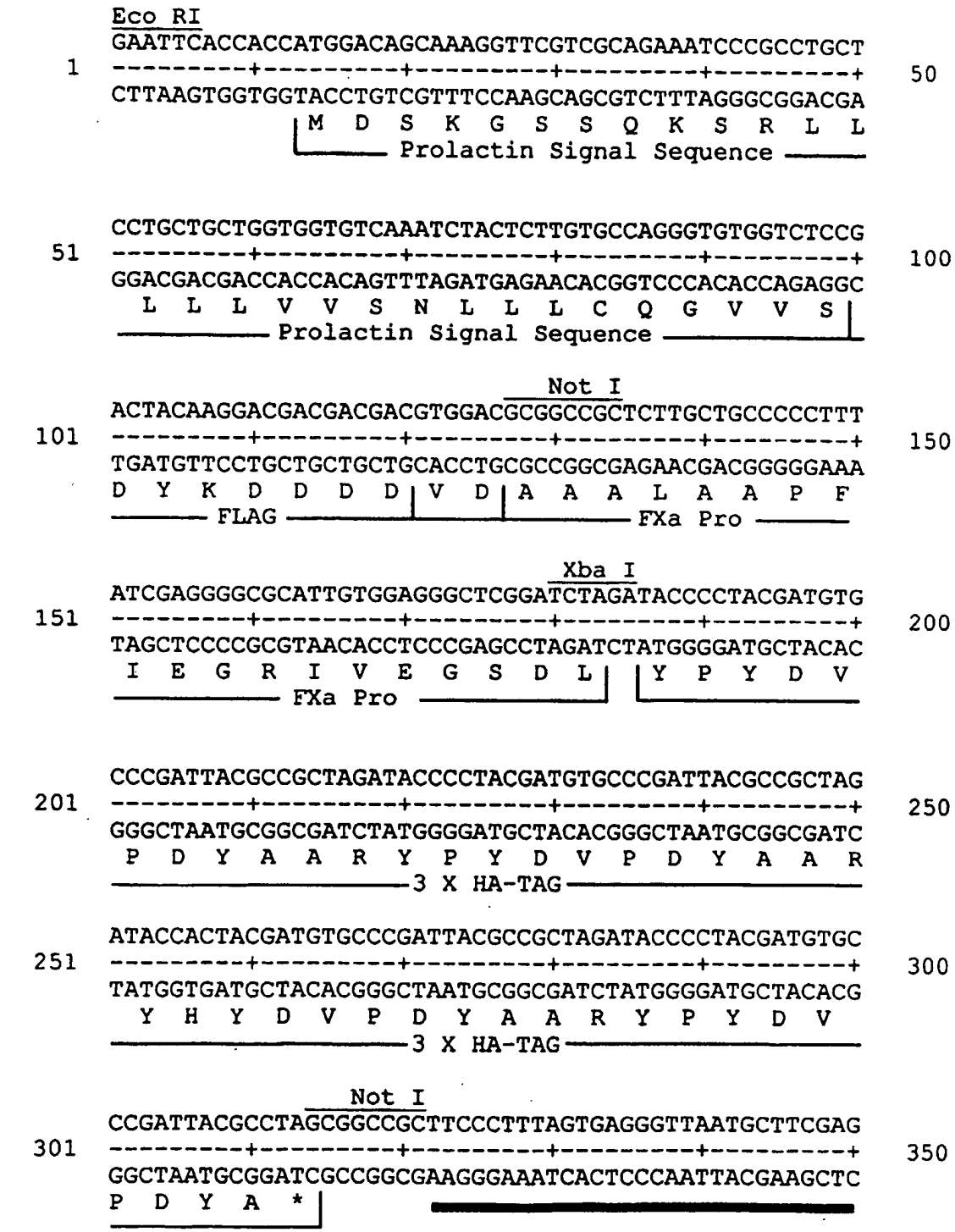
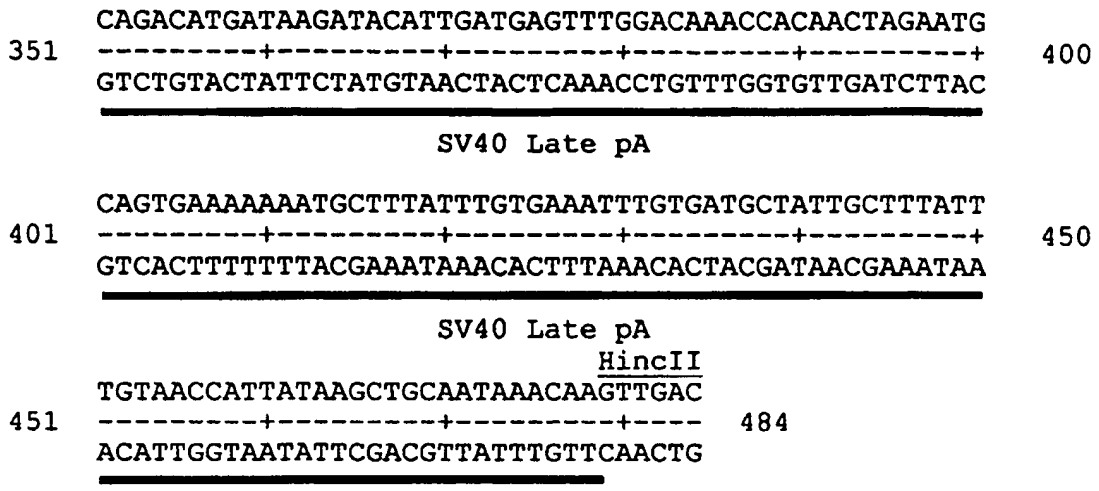


FIG. 2(D)


CAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATG
351
400
GTCTGTACTATTCTATGTAACTACTCAAACCTGTTTGGTGTTGATCTTAC
SV40 Late pA
CAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATT
401
450
GTCACTTTTTTTACGAAATAAACACTTTAAACACTACGATAACGAAATAA
SV40 Late pA
HincII
TGTAACCATTATAAGCTGCAATAAACAAGTTGAC
451
484
ACATTGGTAATATTCGACGTTATTTGTTCAACTG

SEQ.ID.NO.:4

FIG. 2(E)

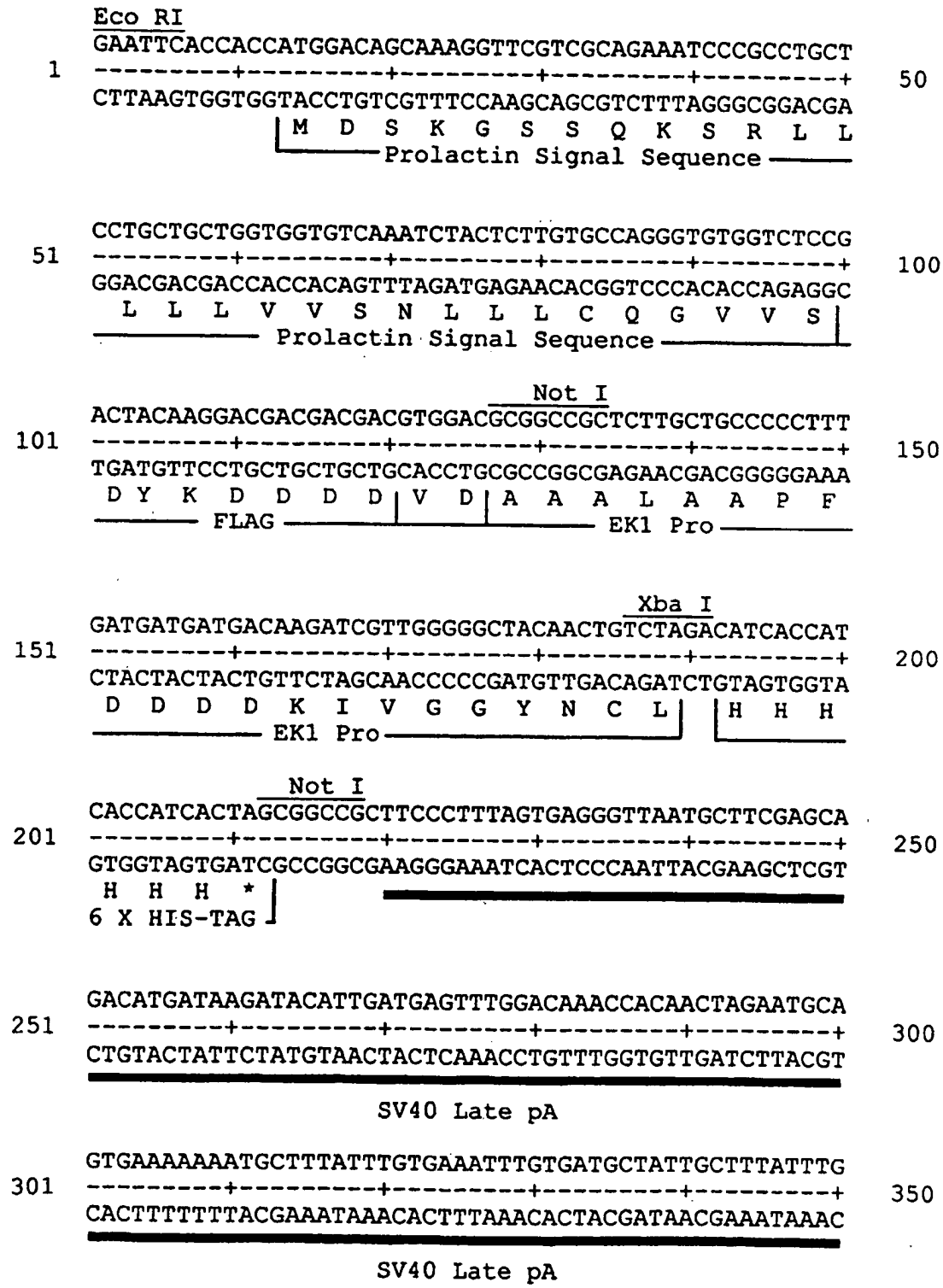


FIG. 2(F)

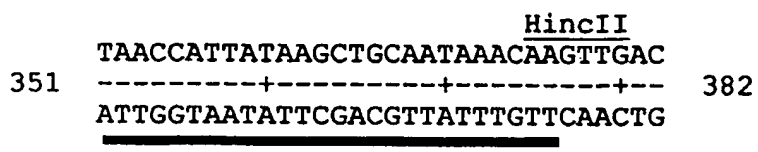

HincII
TAACCATTATAAGCTGCAATAAACAAGTTGAC
351
382
ATTGGTAATATTCGACGTTATTTGTTCAACTG

FIG. 2(G)

SEQ.ID.NO.:5

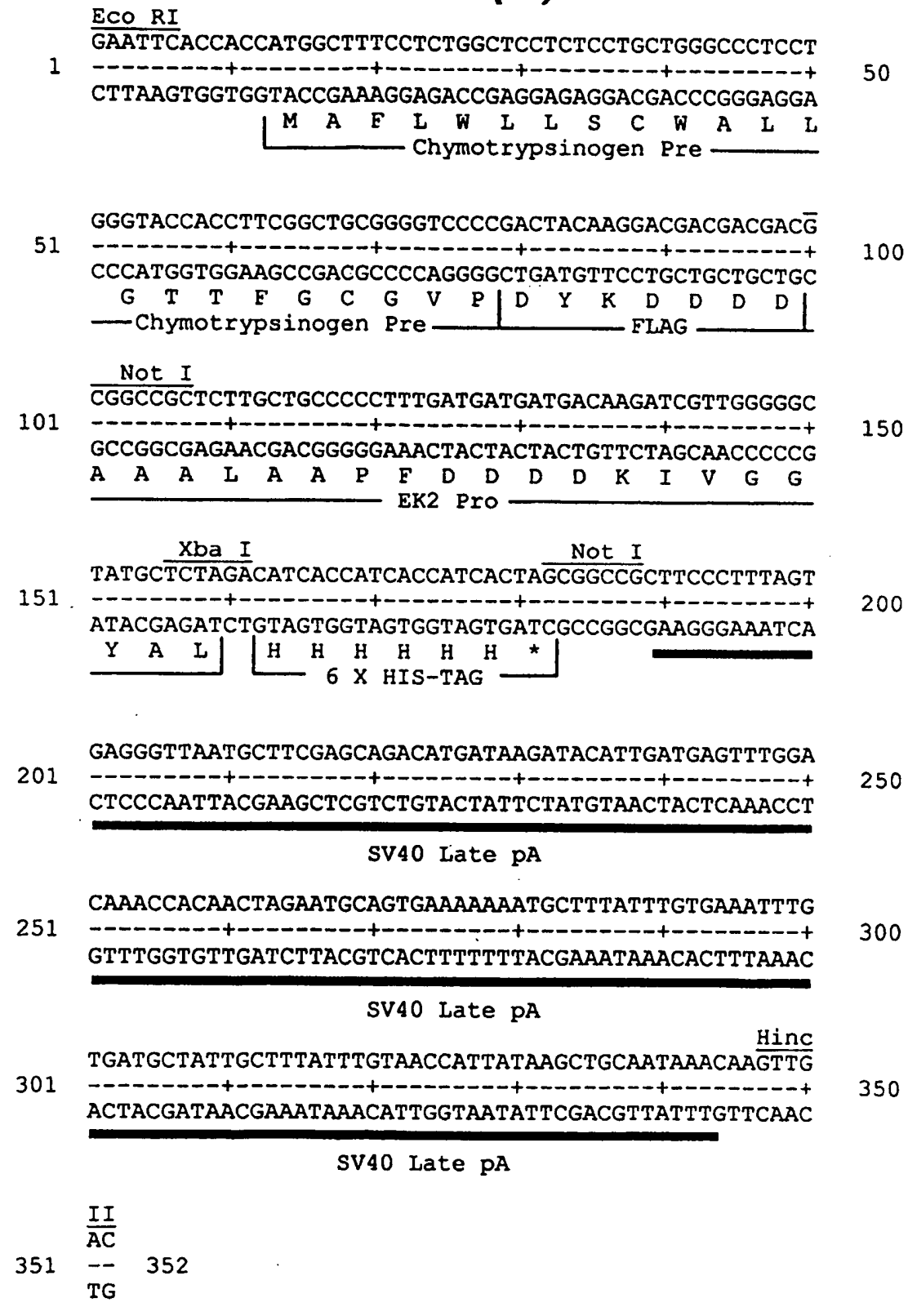


FIG. 2(H)

SEQ.ID.NO.:6

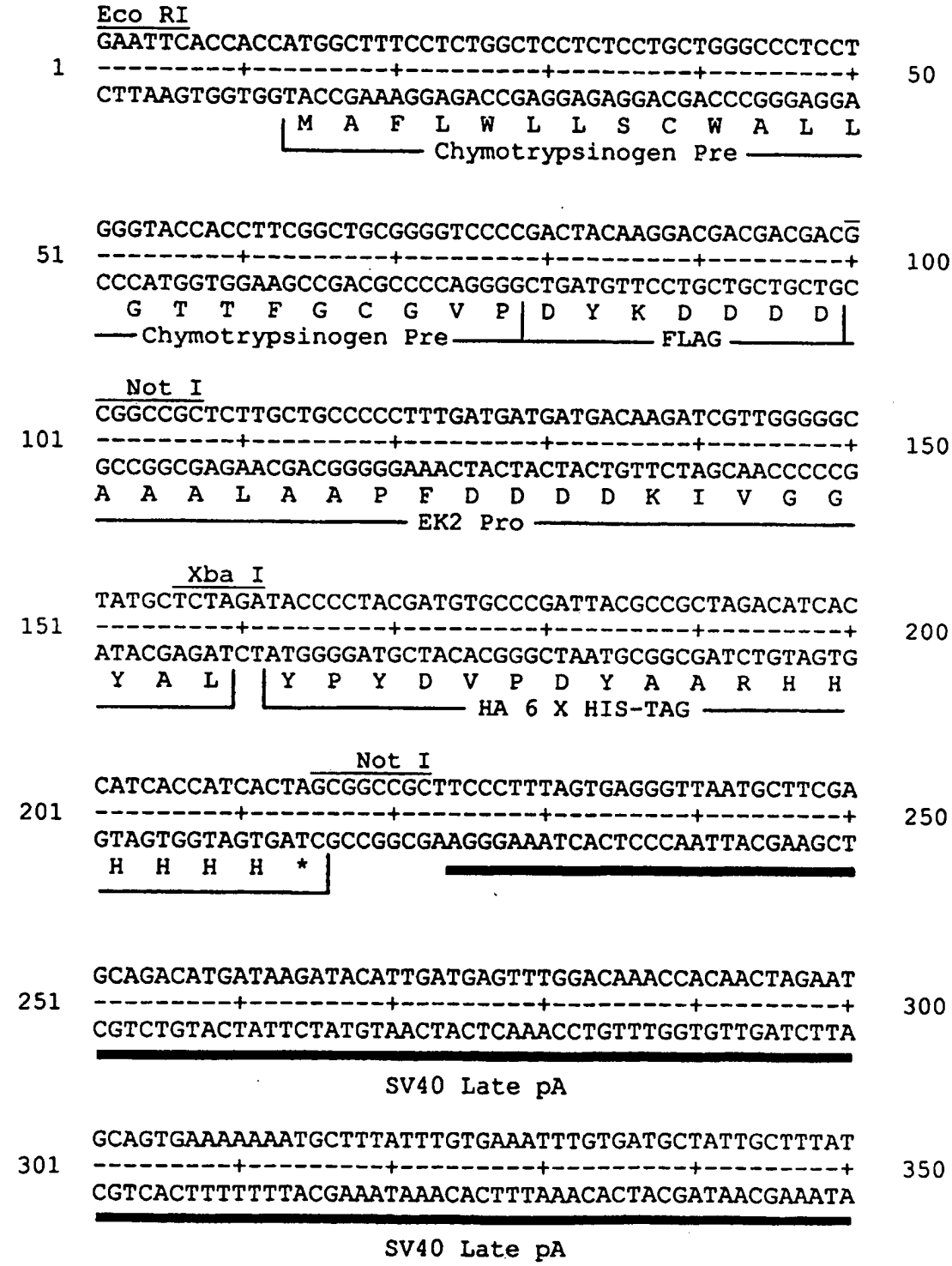
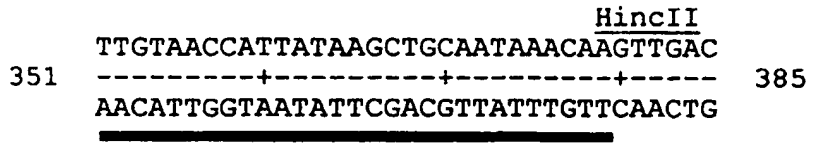


FIG. 2(I)


HincII
TTGTAACCATTATAAGCTGCAATAAACAAGTTGAC
351
385
AACATTGGTAATATTCGACGTTATTTGTTCAACTG

SEQ.ID.NO.:7

FIG. 3(A)

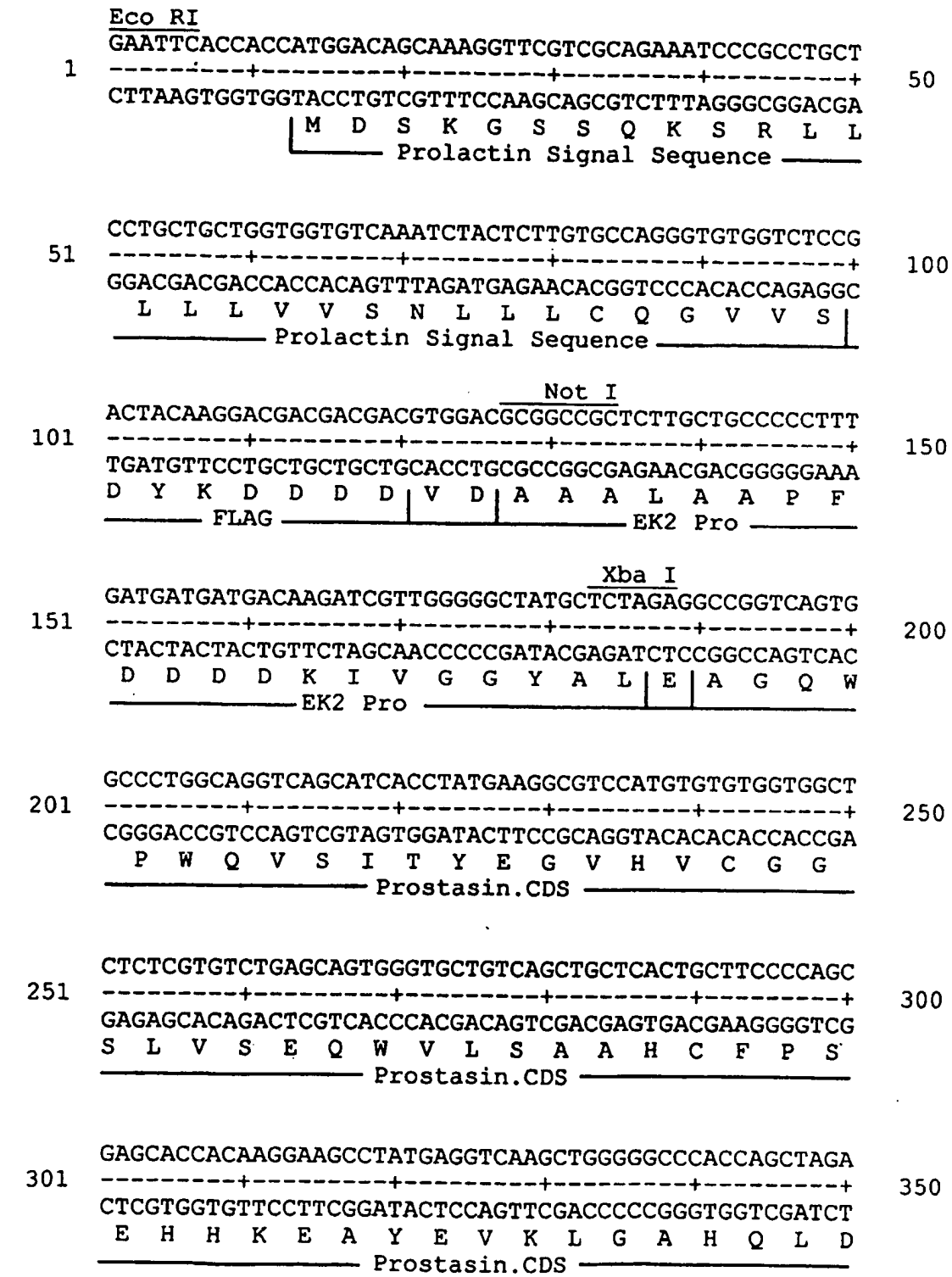


FIG. 3(B)

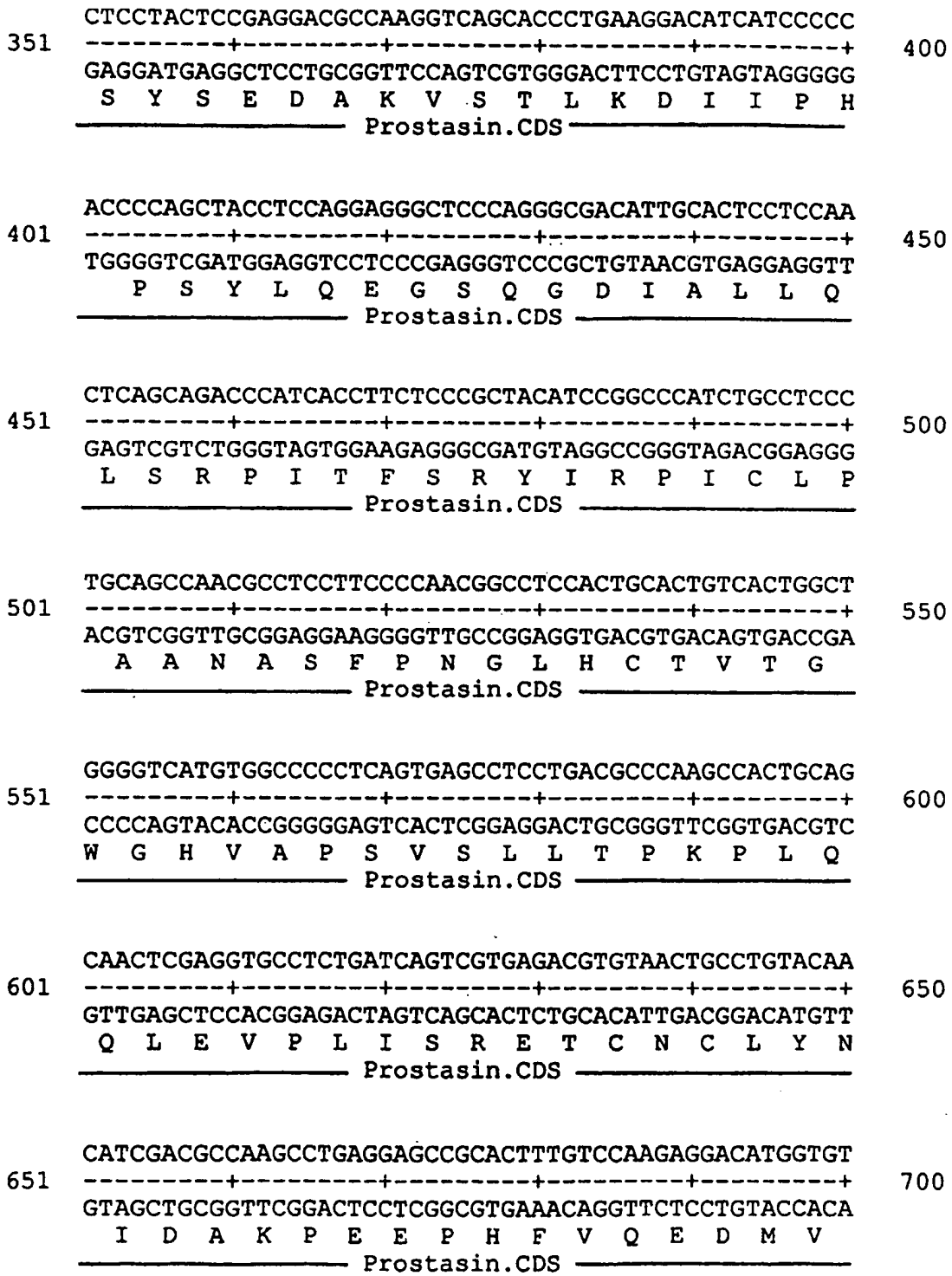

CTCCTACTCCGAGGACGCCAAGGTCAGCACCCTGAAGGACATCATCCCCC
351
GAGGATGAGGCTCCTGCGGTTCCAGTCGTGGGACTTCCTGTAGTAGGGGG
400
S Y S E D A K V S T L K D I I P H
Prostasin.CDS
ACCCCAGCTACCTCCAGGAGGGCTCCCAGGGCGACATTGCACTCCTCCAA
401
TGGGGTCGATGGAGGTCCTCCCGAGGGTCCCGCTGTAACGTGAGGAGGTT
450
P S Y L Q E G S Q G D I A L L Q
Prostasin.CDS
CTCAGCAGACCCATCACCTTCTCCCGCTACATCCGGCCCATCTGCCTCCC
451
GAGTCGTCTGGGTAGTGGAAGAGGGCGATGTAGGCCGGGTAGACGGAGGG
500
L S R P I T F S R Y I R P I C L P
Prostasin.CDS
TGCAGCCAACGCCTCCTTCCCCAACGGCCTCCACTGCACTGTCACTGGCT
501
ACGTCGGTTGCGGAGGAAGGGGTTGCCGGAGGTGACGTGACAGTGACCGA
550
A A N A S F P N G L H C T V T G
Prostasin.CDS
GGGGTCATGTGGCCCCCTCAGTGAGCCTCCTGACGCCCAAGCCACTGCAG
551
CCCCAGTACACCGGGGGAGTCACTCGGAGGACTGCGGGTTCGGTGACGTC
600
W G H V A P S V S L L T P K P L Q
Prostasin.CDS
CAACTCGAGGTGCCTCTGATCAGTCGTGAGACGTGTAACTGCCTGTACAA
601
GTTGAGCTCCACGGAGACTAGTCAGCACTCTGCACATTGACGGACATGTT
650
Q L E V P L I S R E T C N C L Y N
Prostasin.CDS
CATCGACGCCAAGCCTGAGGAGCCGCACTTTGTCCAAGAGGACATGGTGT
651
GTAGCTGCGGTTCGGACTCCTCGGCGTGAAACAGGTTCTCCTGTACCACA
700
I D A K P E E P H F V Q E D M V
Prostasin.CDS

FIG. 3(C)

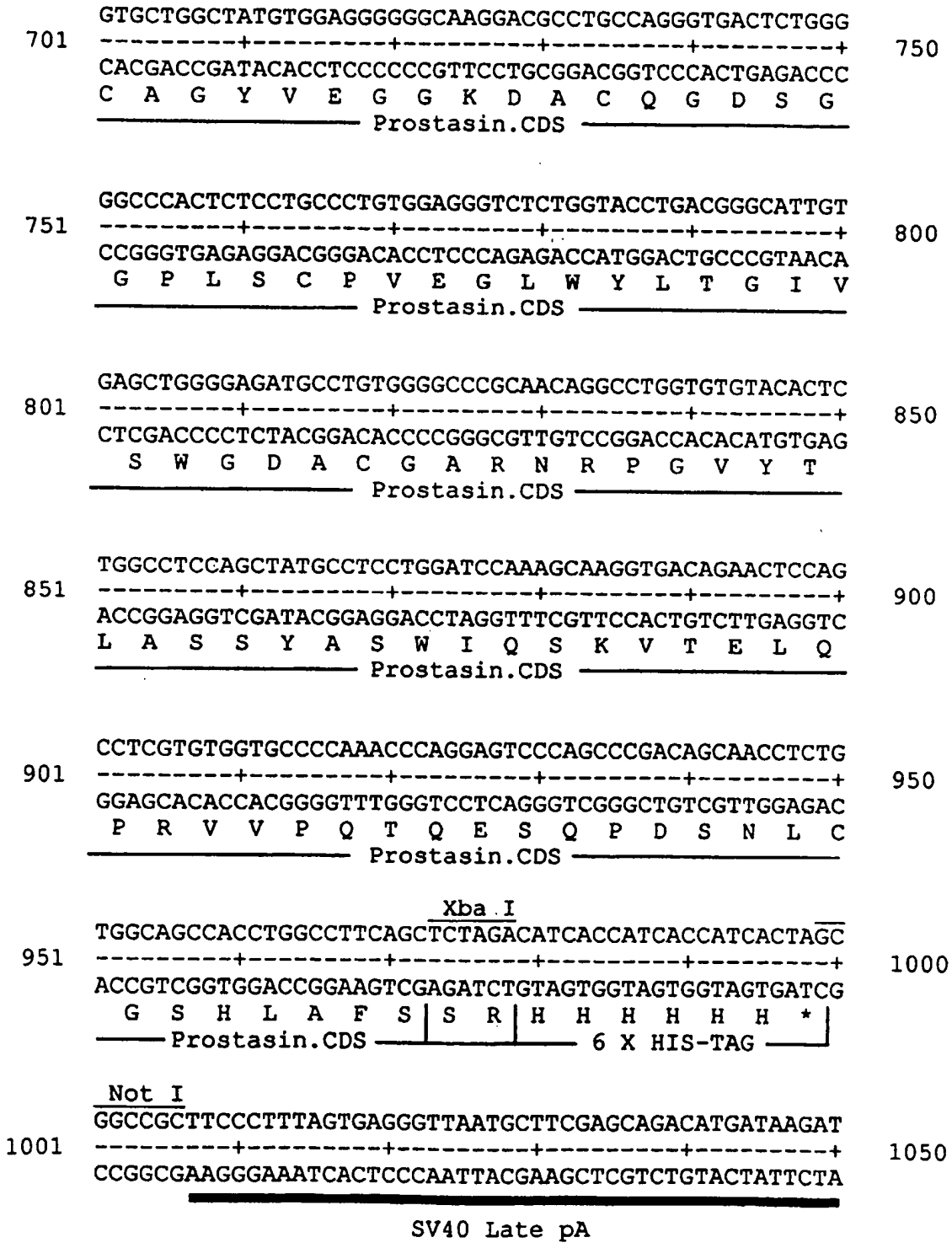

GTGCTGGCTATGTGGAGGGGGGCAAGGACGCCTGCCAGGGTGACTCTGGG
701
750
CACGACCGATACACCTCCCCCCGTTCCTGCGGACGGTCCCACTGAGACCC
C A G Y V E G G K D A C Q G D S G
Prostasin.CDS
GGCCCACTCTCCTGCCCTGTGGAGGGTCTCTGGTACCTGACGGGCATTGT
751
800
CCGGGTGAGAGGACGGGACACCTCCCAGAGACCATGGACTGCCCGTAACA
G P L S C P V E G L W Y L T G I V
Prostasin.CDS
GAGCTGGGGAGATGCCTGTGGGGCCCGCAACAGGCCTGGTGTGTACACTC
801
850
CTCGACCCCTCTACGGACACCCCGGGCGTTGTCCGGACCACACATGTGAG
S W G D A C G A R N R P G V Y T
Prostasin.CDS
TGGCCTCCAGCTATGCCTCCTGGATCCAAAGCAAGGTGACAGAACTCCAG
851
900
ACCGGAGGTCGATACGGAGGACCTAGGTTTCGTTCCACTGTCTTGAGGTC
L A S S Y A S W I Q S K V T E L Q
Prostasin.CDS
CCTCGTGTGGTGCCCCAAACCCAGGAGTCCCAGCCCGACAGCAACCTCTG
901
950
GGAGCACACCACGGGGTTTGGGTCCTCAGGGTCGGGCTGTCGTTGGAGAC
P R V V P Q T Q E S Q P D S N L C
Prostasin.CDS
Xba I
TGGCAGCCACCTGGCCTTCAGCTCTAGACATCACCATCACCATCACTAGC
951
1000
ACCGTCGGTGGACCGGAAGTCGAGATCTGTAGTGGTAGTGGTAGTGATCG
G S H L A F S S R H H H H H H *
Prostasin.CDS
6 X HIS-TAG
Not I
GGCCGCTTCCCTTTAGTGAGGGTTAATGCTTCGAGCAGACATGATAAGAT
1001
1050
CCGGCGAAGGGAAATCACTCCCAATTACGAAGCTCGTCTGTACTATTCTA
SV40 Late pA

FIG. 4(A)

SEQ.ID.NO.:8

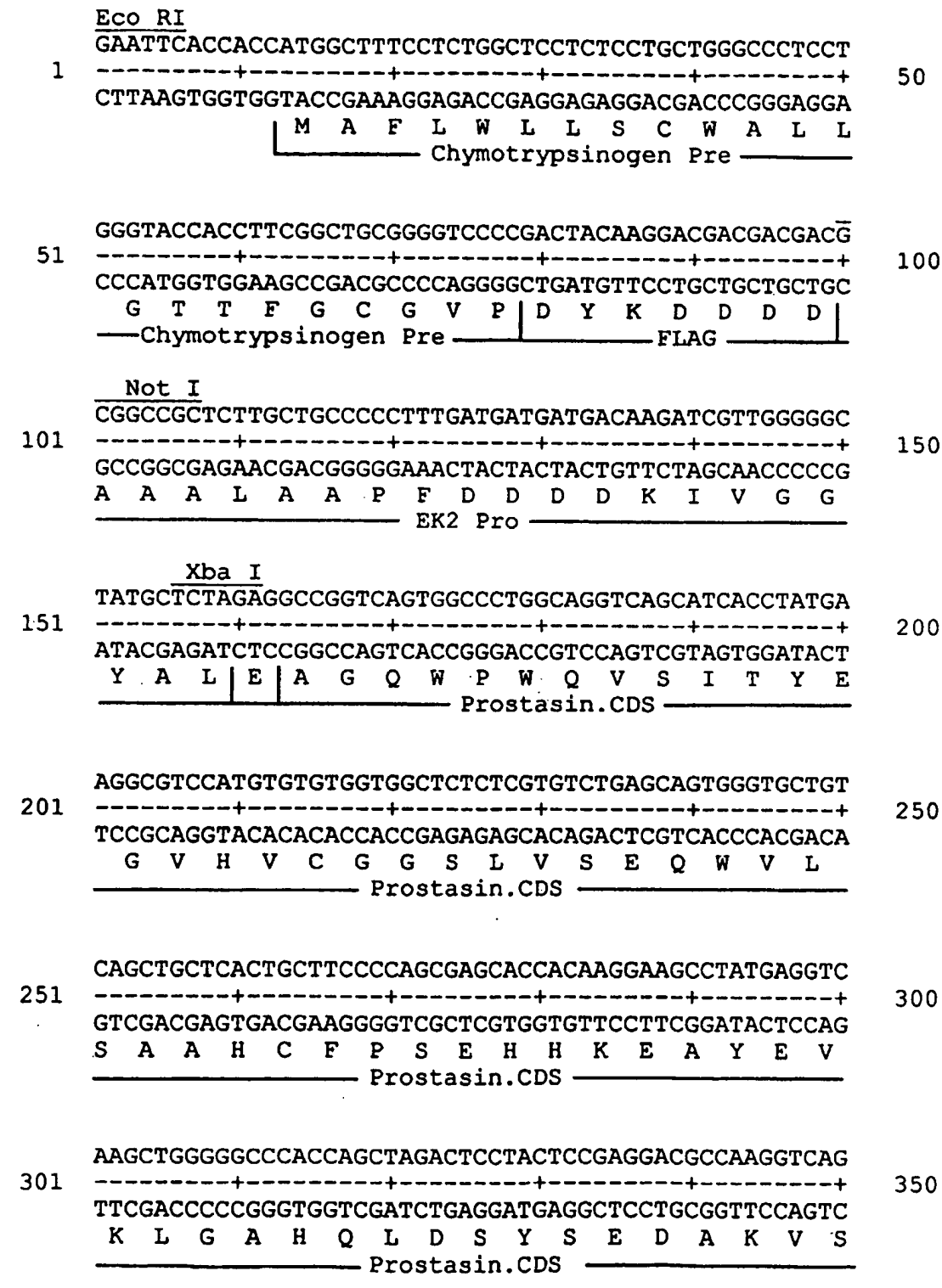


FIG. 4(B)

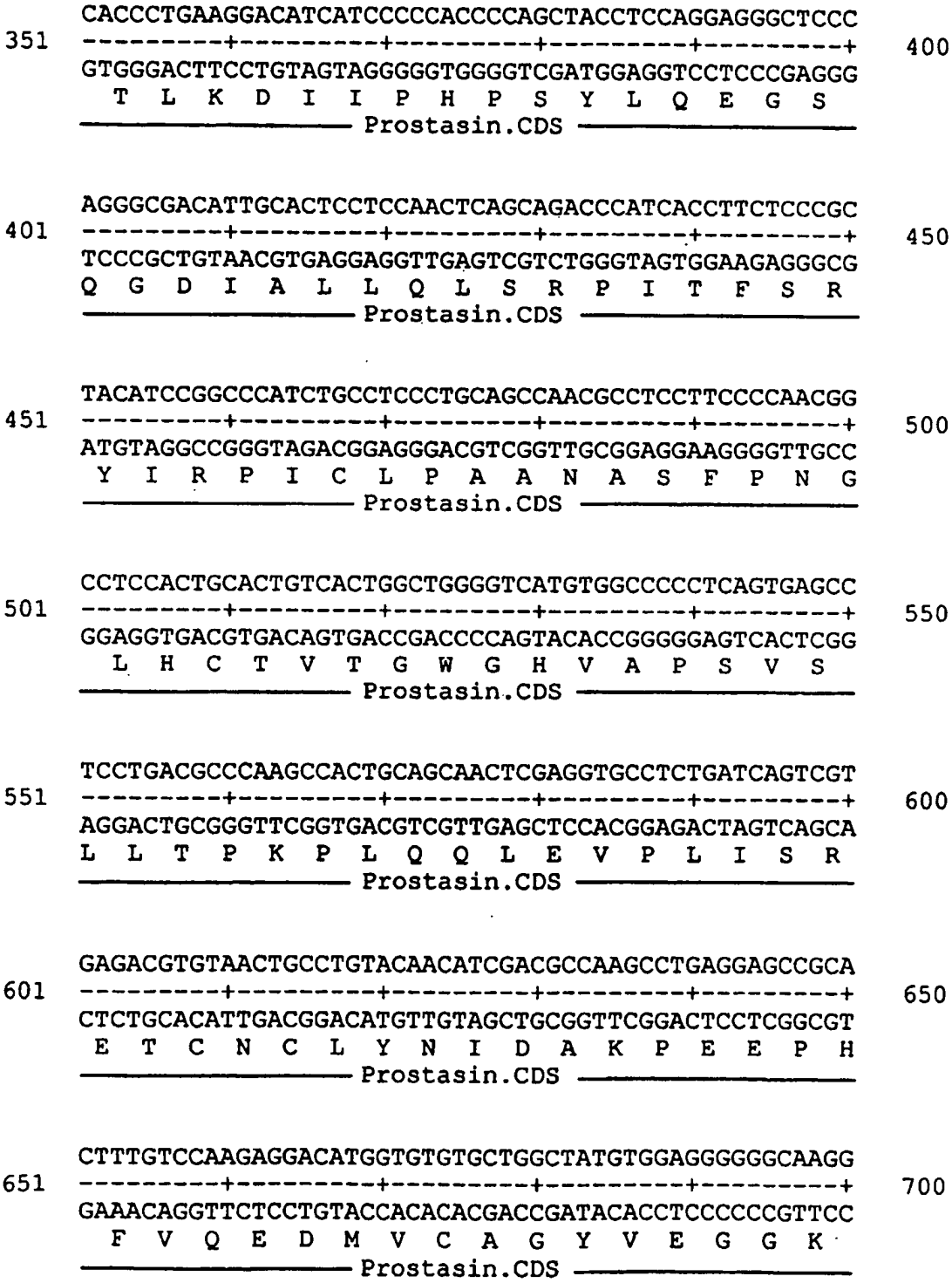

CACCCTGAAGGACATCATCCCCCACCCCAGCTACCTCCAGGAGGGCTCCC
351
400
GTGGGACTTCCTGTAGTAGGGGGTGGGGTCGATGGAGGTCCTCCCGAGGG
T L K D I I P H P S Y L Q E G S
Prostasin.CDS
AGGGCGACATTGCACTCCTCCAACTCAGCAGACCCATCACCTTCTCCCGC
401
450
TCCCGCTGTAACGTGAGGAGGTTGAGTCGTCTGGGTAGTGGAAGAGGGCG
Q G D I A L L Q L S R P I T F S R
Prostasin.CDS
TACATCCGGCCCATCTGCCTCCCTGCAGCCAACGCCTCCTTCCCCAACGG
451
500
ATGTAGGCCGGGTAGACGGAGGGACGTCGGTTGCGGAGGAAGGGGTTGCC
Y I R P I C L P A A N A S F P N G
Prostasin.CDS
CCTCCACTGCACTGTCACTGGCTGGGGTCATGTGGCCCCCTCAGTGAGCC
501
550
GGAGGTGACGTGACAGTGACCGACCCCAGTACACCGGGGGAGTCACTCGG
L H C T V T G W G H V A P S V S
Prostasin.CDS
TCCTGACGCCCAAGCCACTGCAGCAACTCGAGGTGCCTCTGATCAGTCGT
551
600
AGGACTGCGGGTTCGGTGACGTCGTTGAGCTCCACGGAGACTAGTCAGCA
L L T P K P L Q Q L E V P L I S R
Prostasin.CDS
GAGACGTGTAACTGCCTGTACAACATCGACGCCAAGCCTGAGGAGCCGCA
601
650
CTCTGCACATTGACGGACATGTTGTAGCTGCGGTTCGGACTCCTCGGCGT
E T C N C L Y N I D A K P E E P H
Prostasin.CDS
CTTTGTCCAAGAGGACATGGTGTGTGCTGGCTATGTGGAGGGGGGCAAGG
651
700
GAAACAGGTTCTCCTGTACCACACACGACCGATACACCTCCCCCCGTTCC
F V Q E D M V C A G Y V E G G K
Prostasin.CDS

FIG. 4(C)

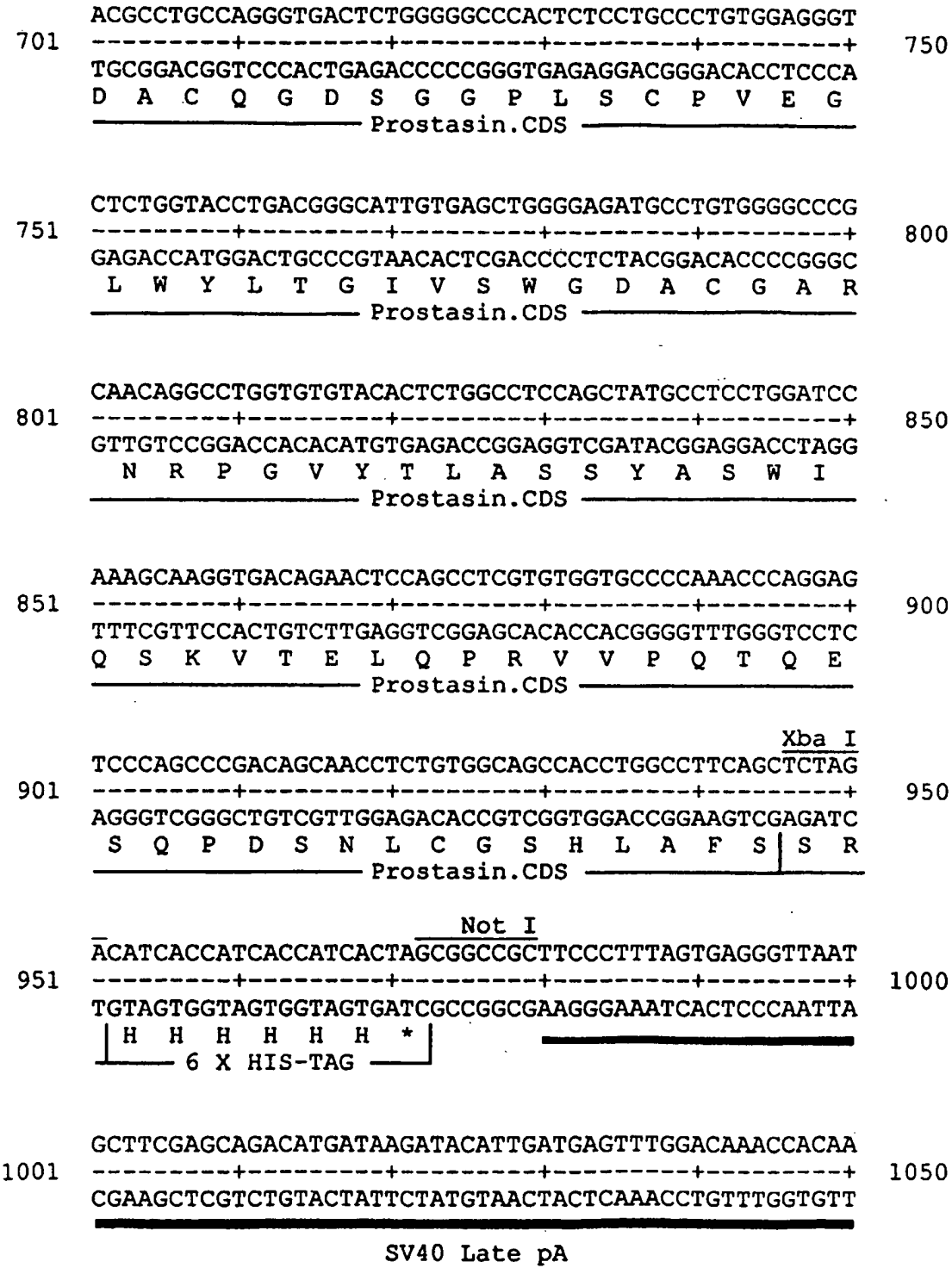
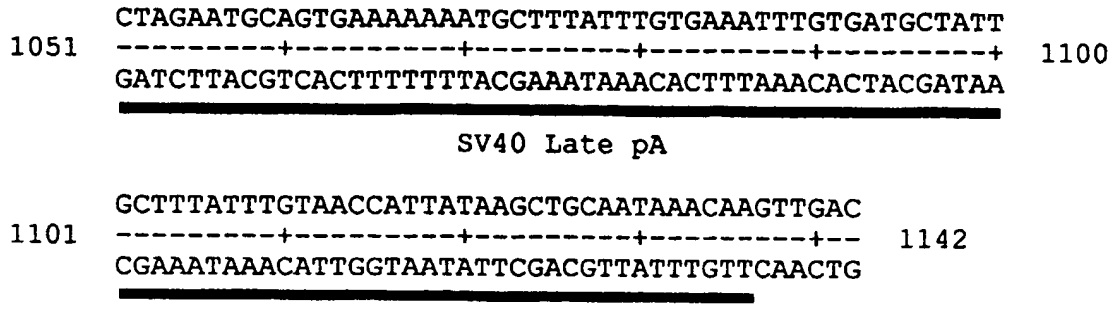

ACGCCTGCCAGGGTGACTCTGGGGGCCCACTCTCCTGCCCTGTGGAGGGT
701
750
TGCGGACGGTCCCACTGAGACCCCCGGGTGAGAGGACGGGACACCTCCCA
D A C Q G D S G G P L S C P V E G
Prostasin.CDS
CTCTGGTACCTGACGGGCATTGTGAGCTGGGGAGATGCCTGTGGGGCCCG
751
800
GAGACCATGGACTGCCCGTAACACTCGACCCCTCTACGGACACCCCGGGC
L W Y L T G I V S W G D A C G A R
Prostasin.CDS
CAACAGGCCTGGTGTGTACACTCTGGCCTCCAGCTATGCCTCCTGGATCC
801
850
GTTGTCCGGACCACACATGTGAGACCGGAGGTCGATACGGAGGACCTAGG
N R P G V Y T L A S S Y A S W I
Prostasin.CDS
AAAGCAAGGTGACAGAACTCCAGCCTCGTGTGGTGCCCCAAACCCAGGAG
851
900
TTTCGTTCCACTGTCTTGAGGTCGGAGCACACCACGGGGTTTGGGTCCTC
Q S K V T E L Q P R V V P Q T Q E
Prostasin.CDS
Xba I
TCCCAGCCCGACAGCAACCTCTGTGGCAGCCACCTGGCCTTCAGCTCTAG
901
950
AGGGTCGGGCTGTCGTTGGAGACACCGTCGGTGGACCGGAAGTCGAGATC
S Q P D S N L C G S H L A F S S R
Prostasin.CDS
Not I
ACATCACCATCACCATCACTAGCGGCCGCTTCCCTTTAGTGAGGGTTAAT
951
1000
TGTAGTGGTAGTGGTAGTGATCGCCGGCGAAGGGAAATCACTCCCAATTA
H H H H H H *
6 X HIS-TAG
GCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAA
1001
1050
CGAAGCTCGTCTGTACTATTCTATGTAACTACTCAAACCTGTTTGGTGTT
SV40 Late pA

FIG. 4(D)


CTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATT
1051
1100
GATCTTACGTCACTTTTTTTACGAAATAAACACTTTAAACACTACGATAA
SV40 Late pA
GCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTGAC
1101
1142
CGAAATAAACATTGGTAATATTCGACGTTATTTGTTCAACTG

SEQ.ID.NO.:9

FIG. 5(A)

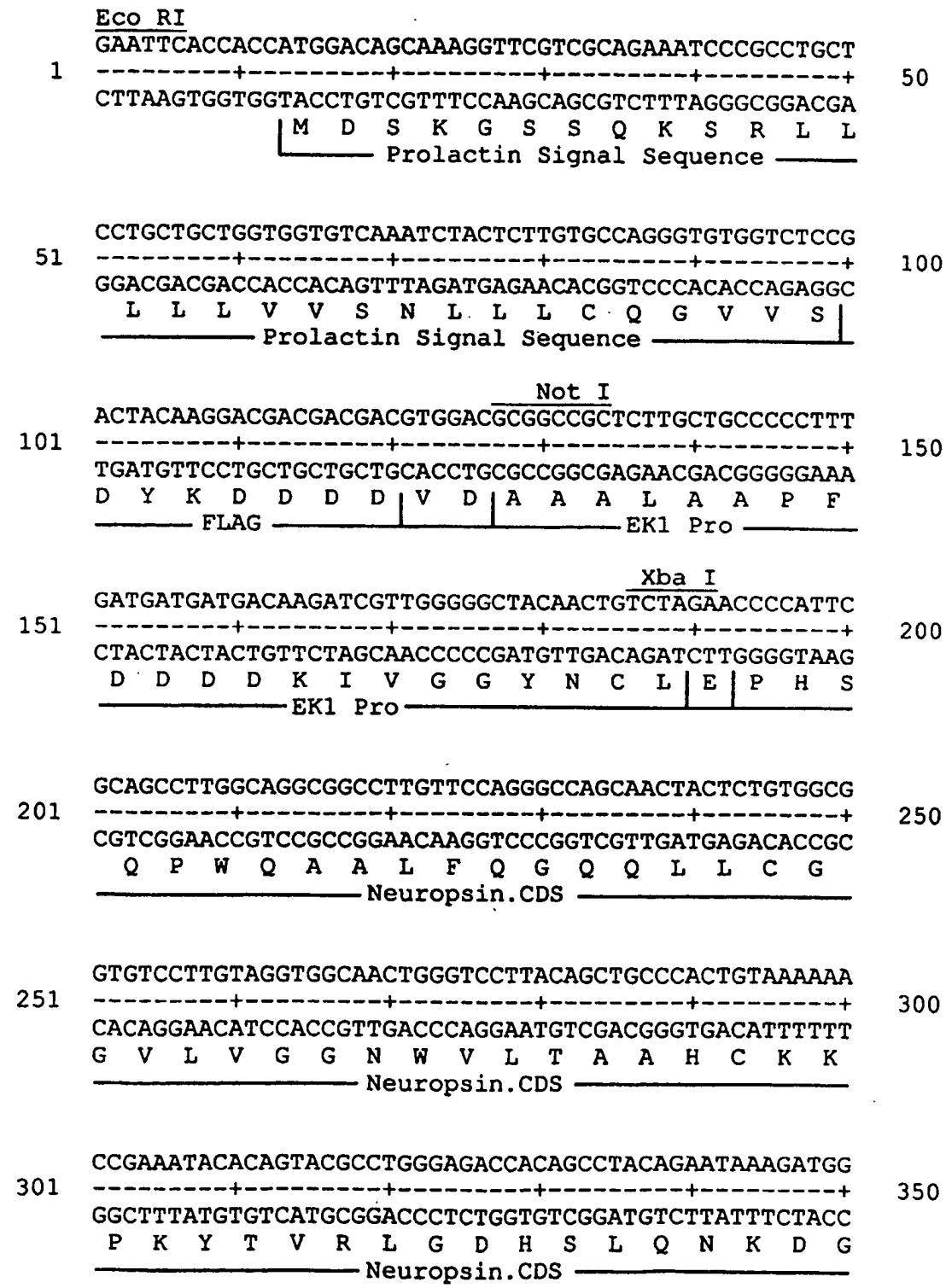


FIG. 5(B)

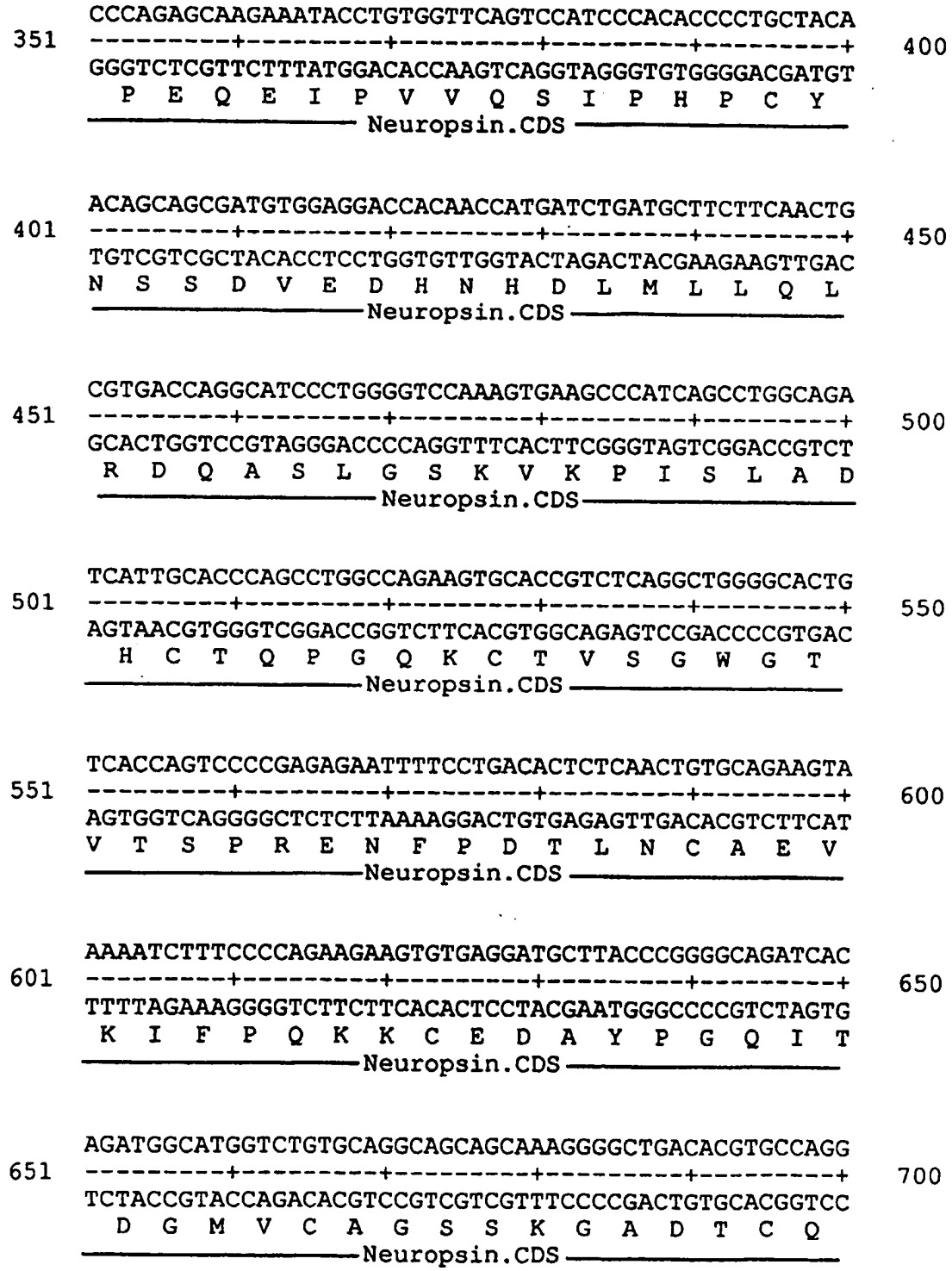

CCCAGAGCAAGAAATACCTGTGGTTCAGTCCATCCCACACCCCTGCTACA
351
400
GGGTCTCGTTCTTTATGGACACCAAGTCAGGTAGGGTGTGGGGACGATGT
P E Q E I P V V Q S I P H P C Y
Neuropsin.CDS
ACAGCAGCGATGTGGAGGACCACAACCATGATCTGATGCTTCTTCAACTG
401
450
TGTCGTCGCTACACCTCCTGGTGTTGGTACTAGACTACGAAGAAGTTGAC
N S S D V E D H N H D L M L L Q L
Neuropsin.CDS
CGTGACCAGGCATCCCTGGGGTCCAAAGTGAAGCCCATCAGCCTGGCAGA
451
500
GCACTGGTCCGTAGGGACCCCAGGTTTCACTTCGGGTAGTCGGACCGTCT
R D Q A S L G S K V K P I S L A D
Neuropsin.CDS
TCATTGCACCCAGCCTGGCCAGAAGTGCACCGTCTCAGGCTGGGGCACTG
501
550
AGTAACGTGGGTCGGACCGGTCTTCACGTGGCAGAGTCCGACCCCGTGAC
H C T Q P G Q K C T V S G W G T
Neuropsin.CDS
TCACCAGTCCCCGAGAGAATTTTCCTGACACTCTCAACTGTGCAGAAGTA
551
600
AGTGGTCAGGGGCTCTCTTAAAAGGACTGTGAGAGTTGACACGTCTTCAT
V T S P R E N F P D T L N C A E V
Neuropsin.CDS
AAAATCTTTCCCCAGAAGAAGTGTGAGGATGCTTACCCGGGGCAGATCAC
601
650
TTTTAGAAAGGGGTCTTCTTCACACTCCTACGAATGGGCCCCGTCTAGTG
K I F P Q K K C E D A Y P G Q I T
Neuropsin.CDS
AGATGGCATGGTCTGTGCAGGCAGCAGCAAAGGGGCTGACACGTGCCAGG
651
700
TCTACCGTACCAGACACGTCCGTCGTCGTTTCCCCGACTGTGCACGGTCC
D G M V C A G S S K G A D T C Q
Neuropsin.CDS

FIG. 5(C)

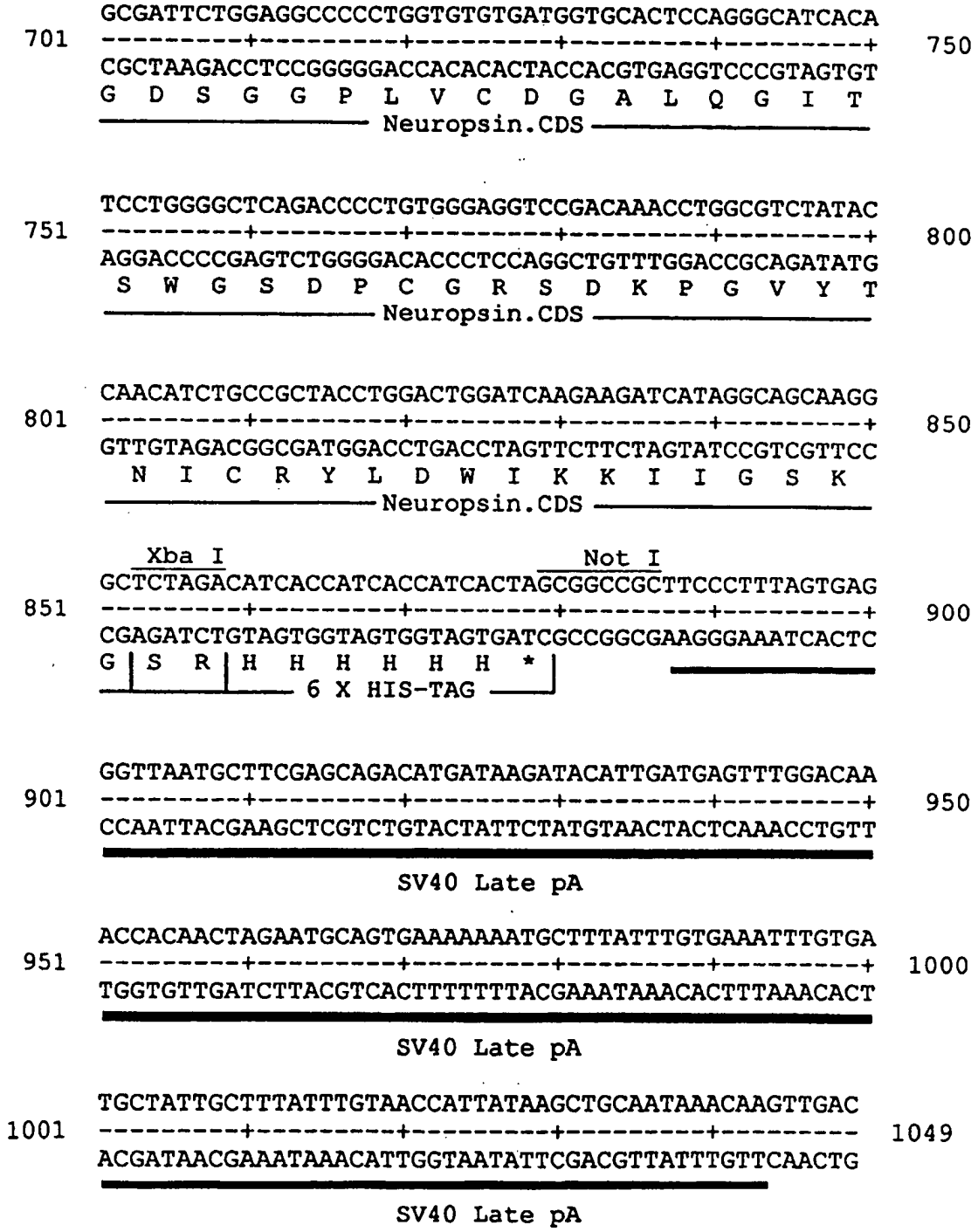

GCGATTCTGGAGGCCCCCTGGTGTGTGATGGTGCACTCCAGGGCATCACA
701
750
CGCTAAGACCTCCGGGGGACCACACACTACCACGTGAGGTCCCGTAGTGT
G D S G G P L V C D G A L Q G I T
Neuropsin.CDS
TCCTGGGGCTCAGACCCCTGTGGGAGGTCCGACAAACCTGGCGTCTATAC
751
800
AGGACCCCGAGTCTGGGGACACCCTCCAGGCTGTTTGGACCGCAGATATG
S W G S D P C G R S D K P G V Y T
Neuropsin.CDS
CAACATCTGCCGCTACCTGGACTGGATCAAGAAGATCATAGGCAGCAAGG
801
850
GTTGTAGACGGCGATGGACCTGACCTAGTTCTTCTAGTATCCGTCGTTCC
N I C R Y L D W I K K I I G S K
Neuropsin.CDS
Xba I
Not I
GCTCTAGACATCACCATCACCATCACTAGCGGCCGCTTCCCTTTAGTGAG
851
900
CGAGATCTGTAGTGGTAGTGGTAGTGATCGCCGGCGAAGGGAAATCACTC
G S R H H H H H H *
6 X HIS-TAG
GGTTAATGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAA
901
950
CCAATTACGAAGCTCGTCTGTACTATTCTATGTAACTACTCAAACCTGTT
SV40 Late pA
ACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGA
951
1000
TGGTGTTGATCTTACGTCACTTTTTTTACGAAATAAACACTTTAAACACT
SV40 Late pA
TGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTGAC
1001
1049
ACGATAACGAAATAAACATTGGTAATATTCGACGTTATTTGTTCAACTG
SV40 Late pA

FIG. 6(A)

SEQ.ID.NO.:10

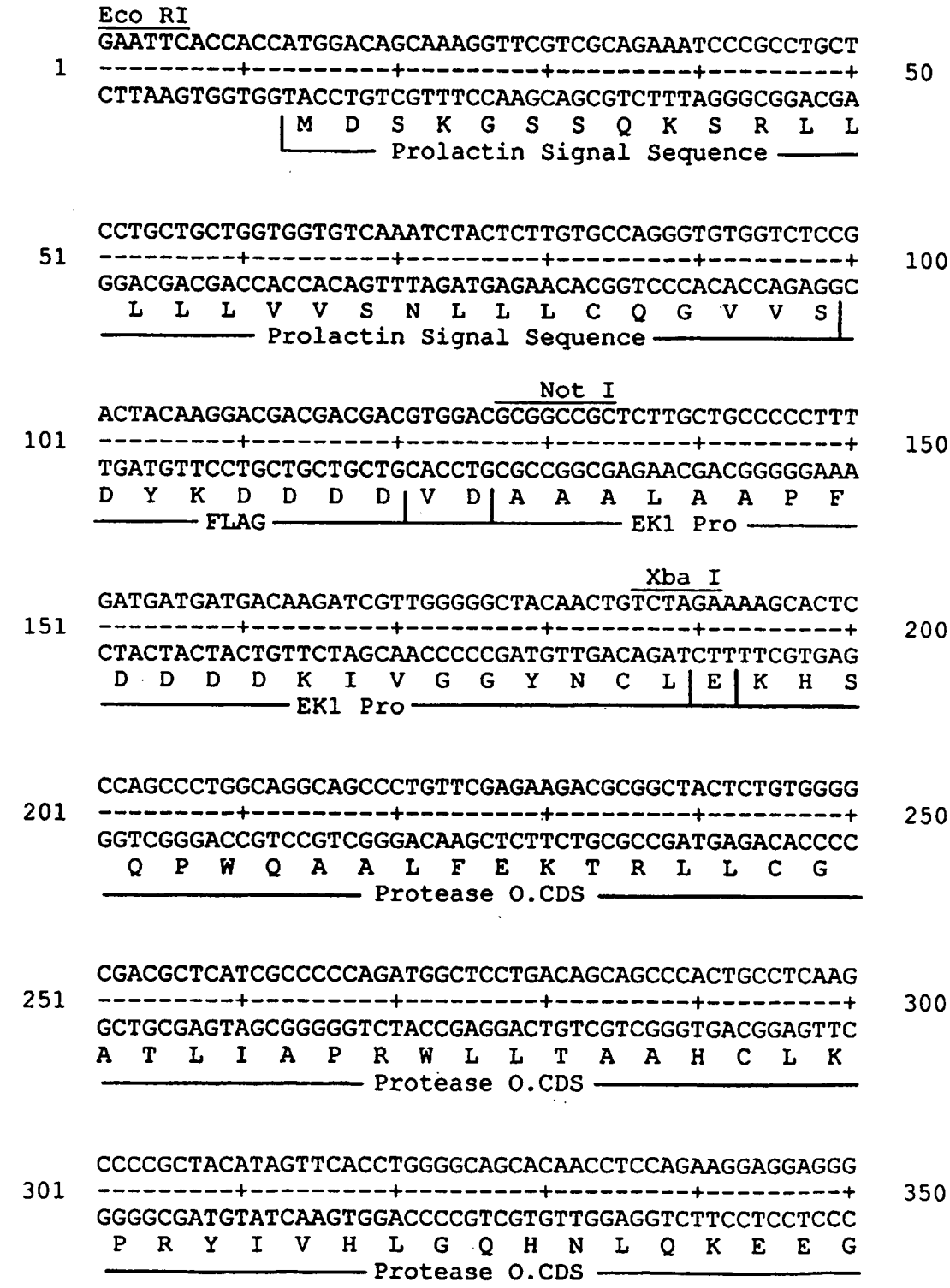


FIG. 6(B)

```
     CTGTGAGCAGACCCGGACAGCCACTGAGTCCTTCCCCCACCCCGGCTTCA
351  ---------+---------+---------+---------+---------+  400
     GACACTCGTCTGGGCCTGTCGGTGACTCAGGAAGGGGGTGGGGCCGAAGT
       C  E  Q  T  R  T  A  T  E  S  F  P  H  P  G  F
     ---------- Protease O.CDS ----------

     ACAACAGCCTCCCCAACAAAGACCACCGCAATGACATCATGCTGGTGAAG
401  ---------+---------+---------+---------+---------+  450
     TGTTGTCGGAGGGGTTGTTTCTGGTGGCGTTACTGTAGTACGACCACTTC
     N  N  S  L  P  N  K  D  H  R  N  D  I  M  L  V  K
     ---------- Protease O.CDS ----------

     ATGGCATCGCCAGTCTCCATCACCTGGGCTGTGCGACCCCTCACCCTCTC
451  ---------+---------+---------+---------+---------+  500
     TACCGTAGCGGTCAGAGGTAGTGGACCCGACACGCTGGGGAGTGGGAGAG
      M  A  S  P  V  S  I  T  W  A  V  R  P  L  T  L  S
     ---------- Protease O.CDS ----------

     CTCACGCTGTGTCACTGCTGGCACCAGCTGCCTCATTTCCGGCTGGGGCA
501  ---------+---------+---------+---------+---------+  550
     GAGTGCGACACAGTGACGACCGTGGTCGACGGAGTAAAGGCCGACCCCGT
       S  R  C  V  T  A  G  T  S  C  L  I  S  G  W  G
     ---------- Protease O.CDS ----------

     GCACGTCCAGCCCCCAGTTACGCCTGCCTCACACCTTGCGATGCGCCAAC
551  ---------+---------+---------+---------+---------+  600
     CGTGCAGGTCGGGGGTCAATGCGGACGGAGTGTGGAACGCTACGCGGTTG
     S  T  S  S  P  Q  L  R  L  P  H  T  L  R  C  A  N
     ---------- Protease O.CDS ----------

     ATCACCATCATTGAGCACCAGAAGTGTGAGAACGCCTACCCCGGCAACAT
601  ---------+---------+---------+---------+---------+  650
     TAGTGGTAGTAACTCGTGGTCTTCACACTCTTGCGGATGGGGCCGTTGTA
      I  T  I  I  E  H  Q  K  C  E  N  A  Y  P  G  N  I
     ---------- Protease O.CDS ----------

     CACAGACACCATGGTGTGTGCCAGCGTGCAGGAAGGGGGCAAGGACTCCT
651  ---------+---------+---------+---------+---------+  700
     GTGTCTGTGGTACCACACACGGTCGCACGTCCTTCCCCCGTTCCTGAGGA
       T  D  T  M  V  C  A  S  V  Q  E  G  G  K  D  S
     ---------- Protease O.CDS ----------
```

FIG. 6(C)

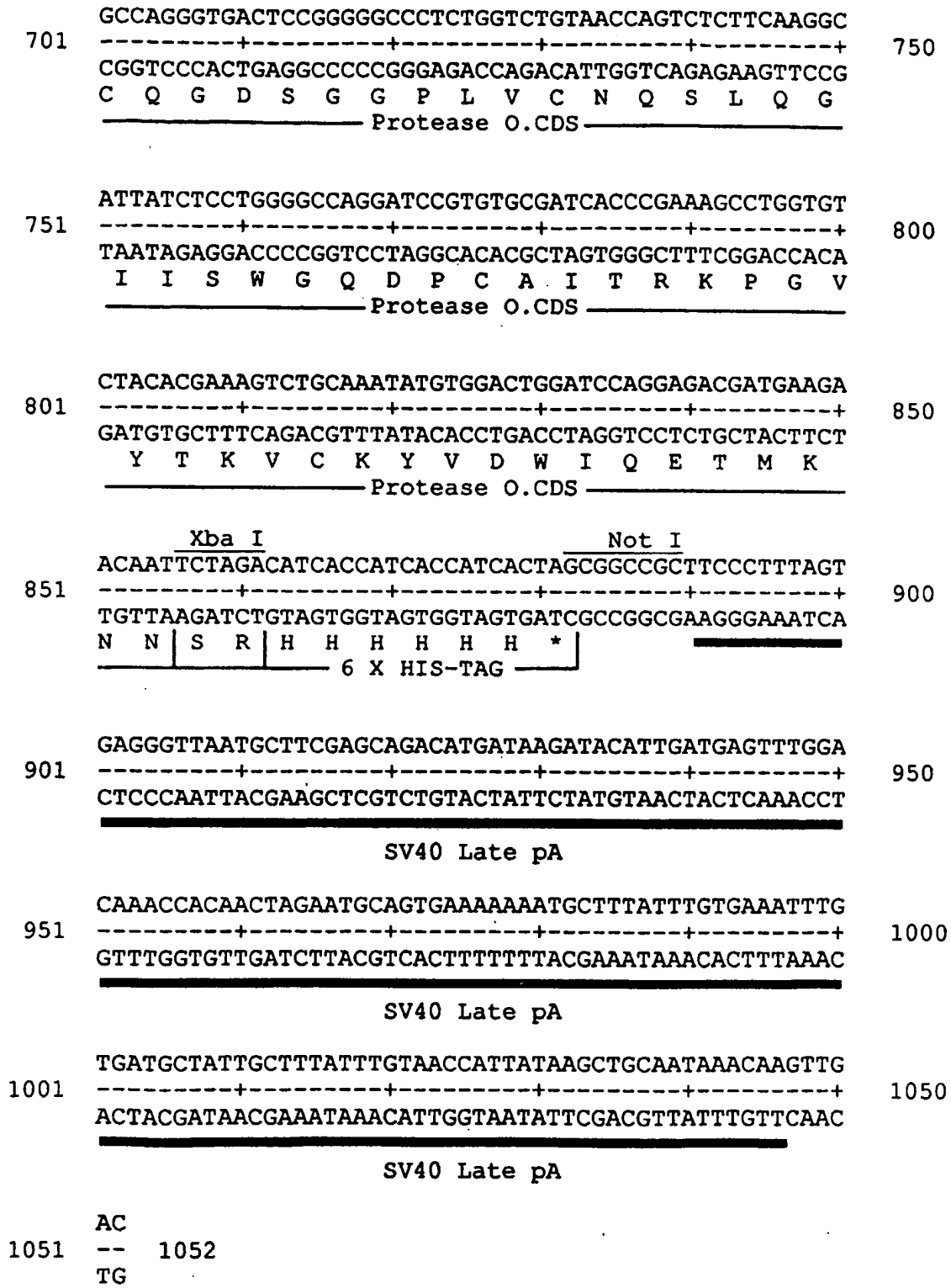
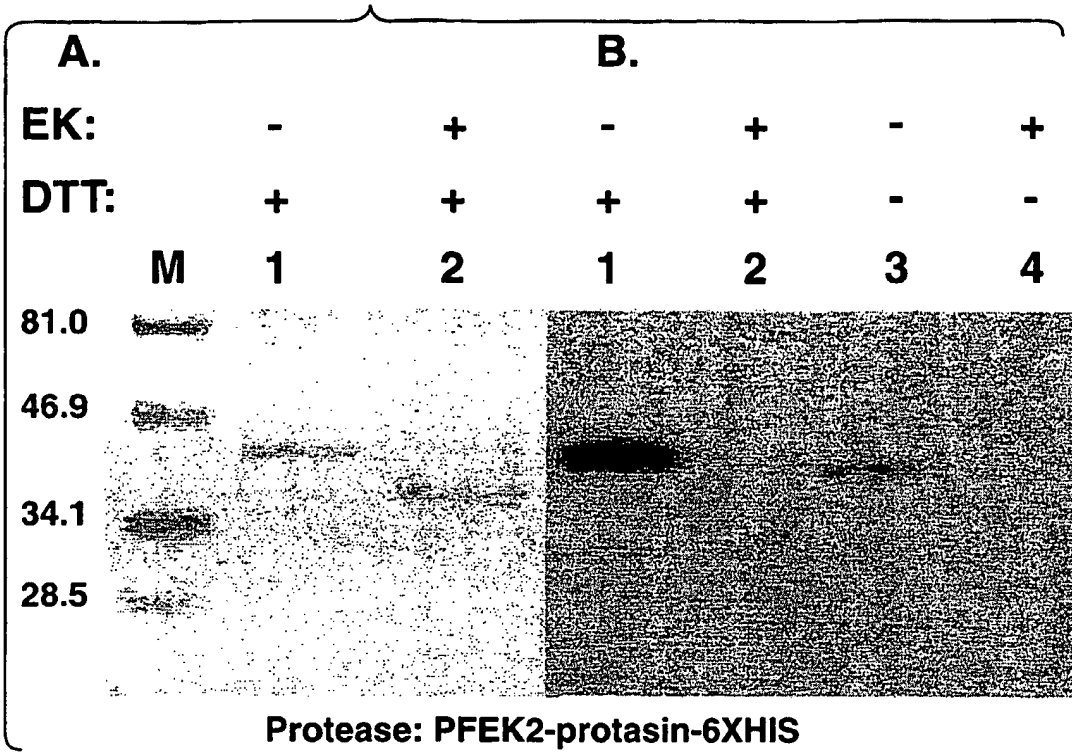
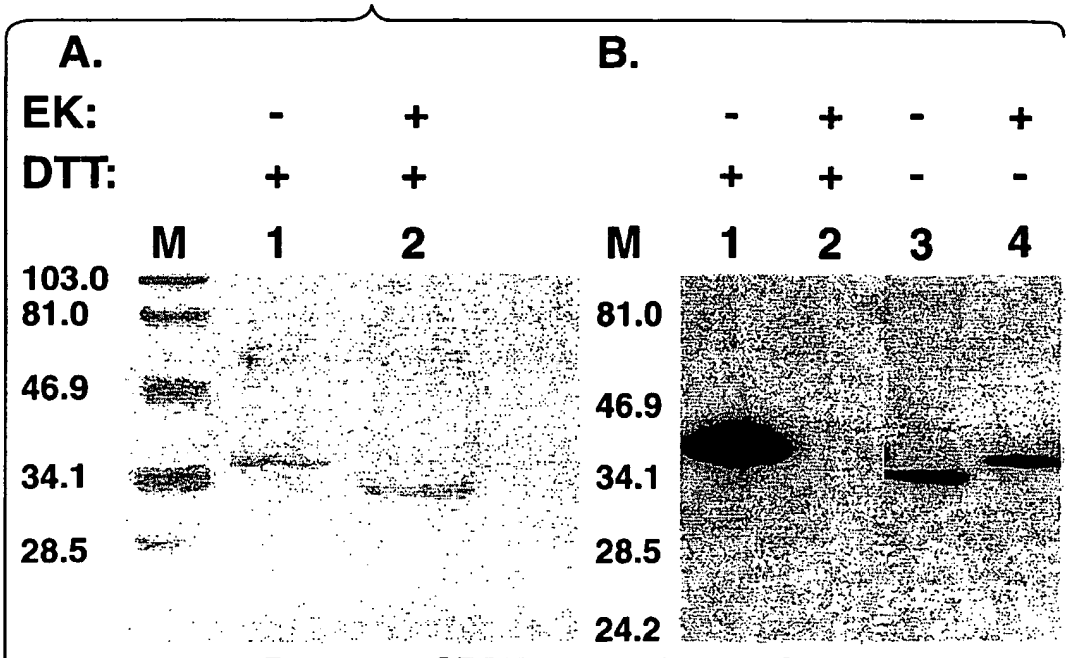

GCCAGGGTGACTCCGGGGGCCCTCTGGTCTGTAACCAGTCTCTTCAAGGC
701
750
CGGTCCCACTGAGGCCCCCGGGAGACCAGACATTGGTCAGAGAAGTTCCG
C Q G D S G G P L V C N Q S L Q G
Protease O.CDS
ATTATCTCCTGGGGCCAGGATCCGTGTGCGATCACCCGAAAGCCTGGTGT
751
800
TAATAGAGGACCCCGGTCCTAGGCACACGCTAGTGGGCTTTCGGACCACA
I I S W G Q D P C A I T R K P G V
Protease O.CDS
CTACACGAAAGTCTGCAAATATGTGGACTGGATCCAGGAGACGATGAAGA
801
850
GATGTGCTTTCAGACGTTTATACACCTGACCTAGGTCCTCTGCTACTTCT
Y T K V C K Y V D W I Q E T M K
Protease O.CDS
Xba I
Not I
ACAATTCTAGACATCACCATCACCATCACTAGCGGCCGCTTCCCTTTAGT
851
900
TGTTAAGATCTGTAGTGGTAGTGGTAGTGATCGCCGGCGAAGGGAAATCA
N N S R H H H H H H *
6 X HIS-TAG
GAGGGTTAATGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGA
901
950
CTCCCAATTACGAAGCTCGTCTGTACTATTCTATGTAACTACTCAAACCT
SV40 Late pA
CAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTG
951
1000
GTTTGGTGTTGATCTTACGTCACTTTTTTTACGAAATAAACACTTTAAAC
SV40 Late pA
TGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTG
1001
1050
ACTACGATAACGAAATAAACATTGGTAATATTCGACGTTATTTGTTCAAC
SV40 Late pA
AC
1051
1052
TG

FIG. 7



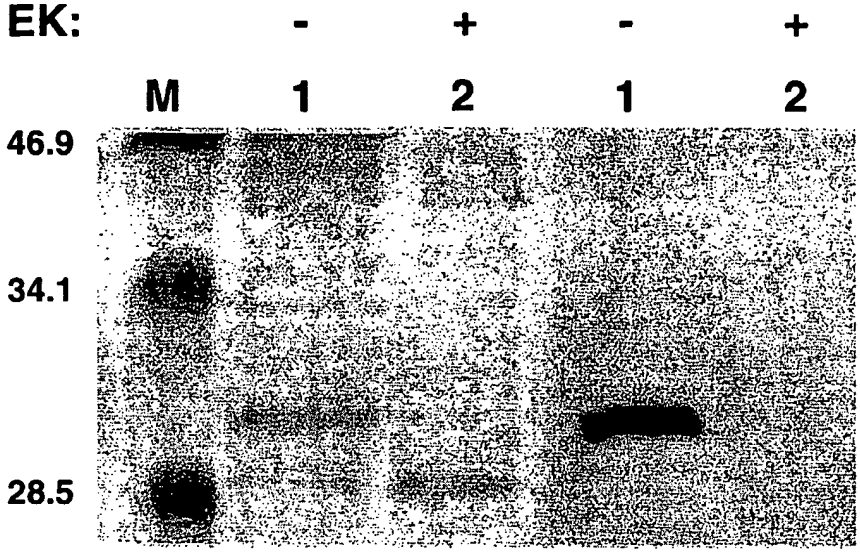
Protease: PFEK2-protasin-6XHIS

FIG. 8



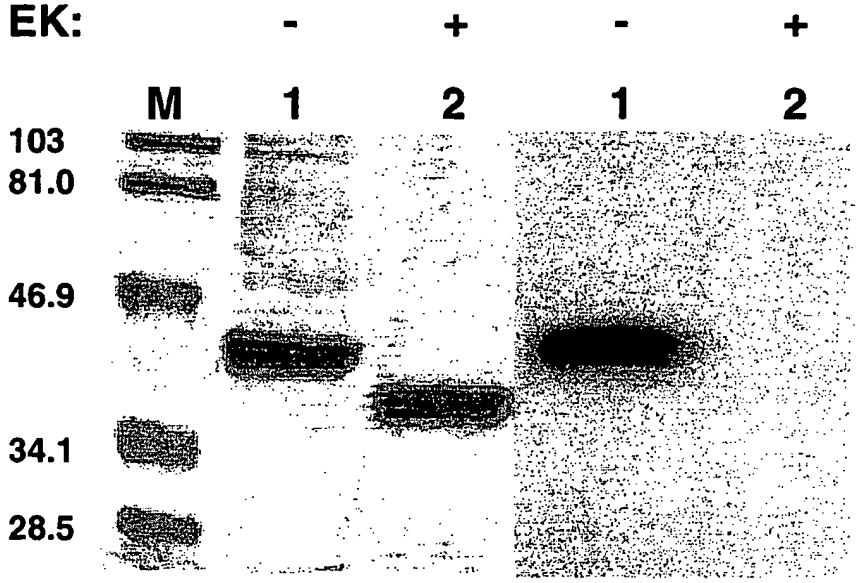
Prot ase: CFEK2-protasin-6XHIS

FIG. 9



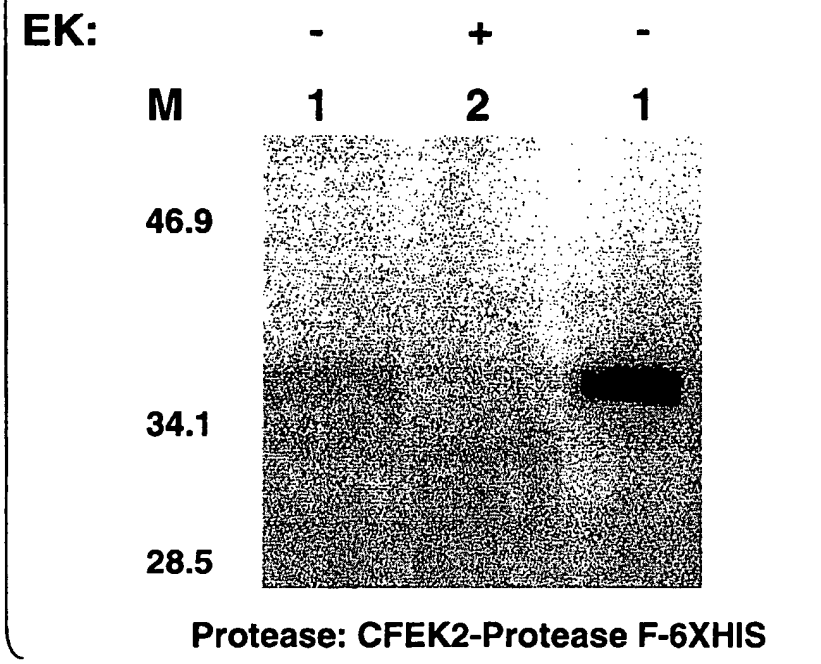
Protease: PFEK1-neuropsin-6XHIS

FIG. 10



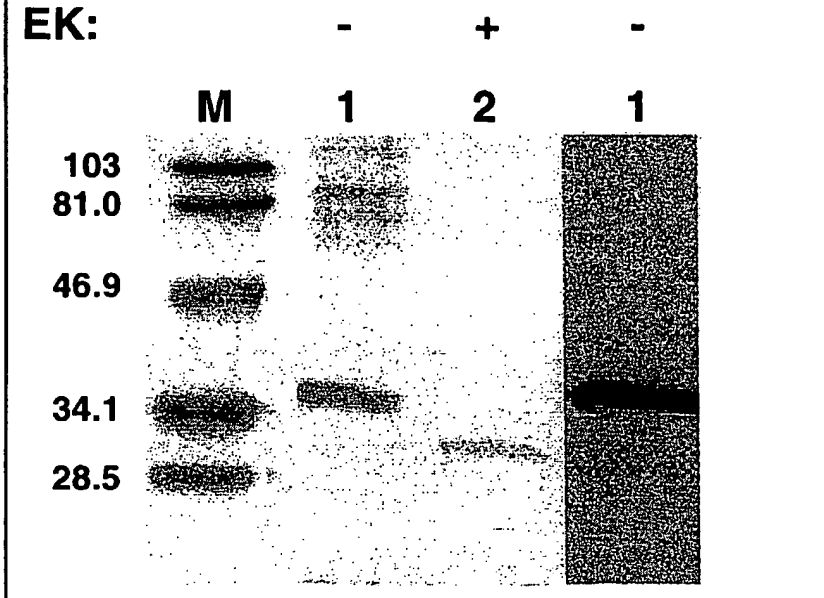
Prot ase: PFEK1-proteas O-6XHIS

FIG. 11



Protease: CFEK2-Protease F-6XHIS

FIG. 12



Protease: PFEK-MH2-6XHIS

SEQ.ID.NO.:53

FIG. 13(A)

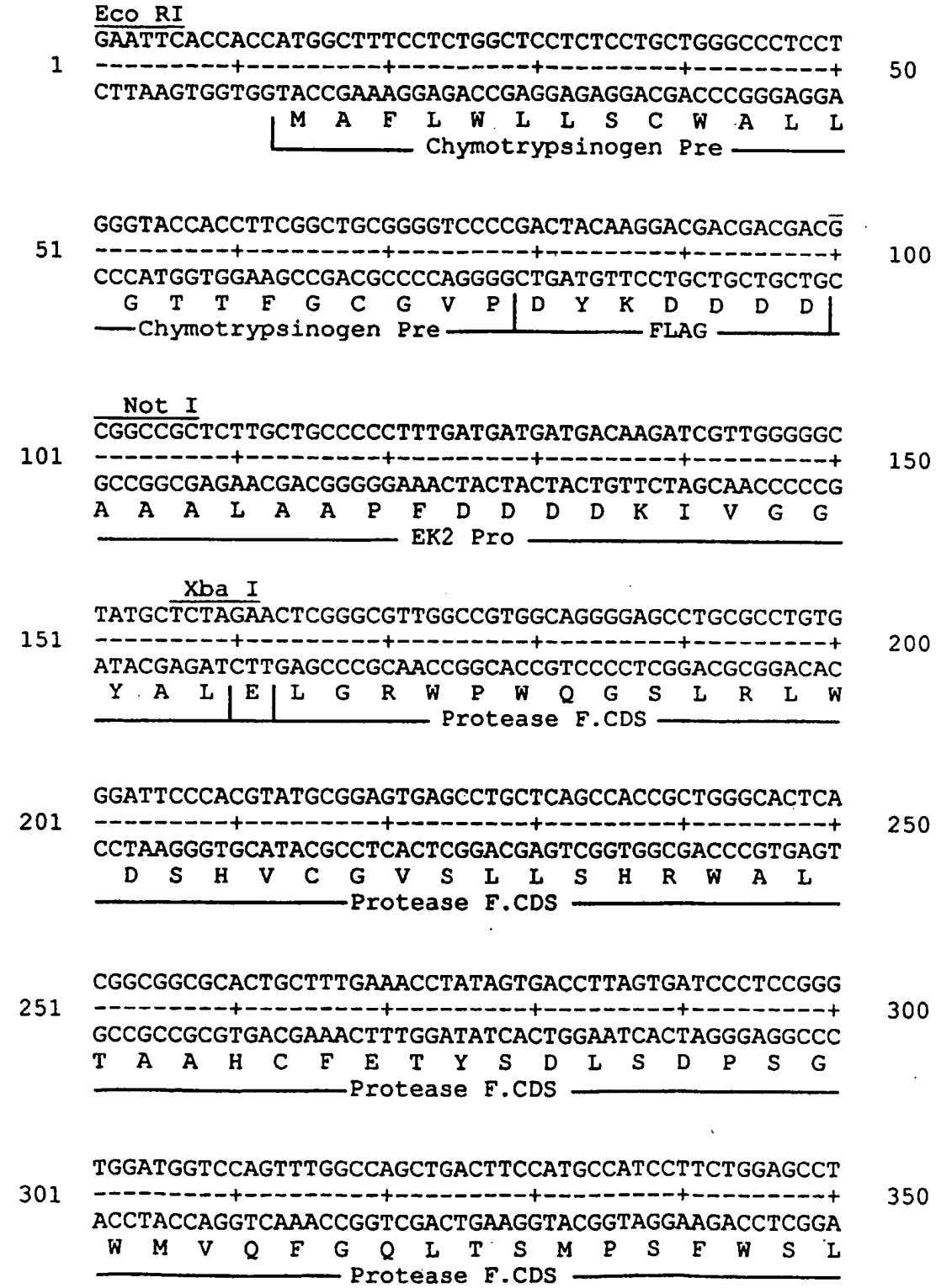


FIG. 13(B)

```
    GCAGGCCTACTACAACCGTTACTTCGTATCGAATATCTATCTGAGCCCTC
351 ---------+---------+---------+---------+---------+ 400
    CGTCCGGATGATGTTGGCAATGAAGCATAGCTTATAGATAGACTCGGGAG
      Q  A  Y  Y  N  R  Y  F  V  S  N  I  Y  L  S  P
    ——————————— Protease F.CDS ———————————

    GCTACCTGGGGAATTCACCCTATGACATTGCCTTGGTGAAGCTGTCTGCA
401 ---------+---------+---------+---------+---------+ 450
    CGATGGACCCCTTAAGTGGGATACTGTAACGGAACCACTTCGACAGACGT
    R  Y  L  G  N  S  P  Y  D  I  A  L  V  K  L  S  A
    ——————————— Protease F.CDS ———————————

    CCTGTCACCTACACTAAACACATCCAGCCCATCTGTCTCCAGGCCTCCAC
451 ---------+---------+---------+---------+---------+ 500
    GGACAGTGGATGTGATTTGTGTAGGTCGGGTAGACAGAGGTCCGGAGGTG
      P  V  T  Y  T  K  H  I  Q  P  I  C  L  Q  A  S  T
    ——————————— Protease F.CDS ———————————

    ATTTGAGTTTGAGAACCGGACAGACTGCTGGGTGACTGGCTGGGGGTACA
501 ---------+---------+---------+---------+---------+ 550
    TAAACTCAAACTCTTGGCCTGTCTGACGACCCACTGACCGACCCCCATGT
      F  E  F  E  N  R  T  D  C  W  V  T  G  W  G  Y
    ——————————— Protease F.CDS ———————————

    TCAAAGAGGATGAGGCACTGCCATCTCCCCACACCCTCCAGGAAGTTCAG
551 ---------+---------+---------+---------+---------+ 600
    AGTTTCTCCTACTCCGTGACGGTAGAGGGGTGTGGGAGGTCCTTCAAGTC
    I  K  E  D  E  A  L  P  S  P  H  T  L  Q  E  V  Q
    ——————————— Protease F.CDS ———————————

    GTCGCCATCATAAACAACTCTATGTGCAACCACCTCTTCCTCAAGTACAG
601 ---------+---------+---------+---------+---------+ 650
    CAGCGGTAGTATTTGTTGAGATACACGTTGGTGGAGAAGGAGTTCATGTC
      V  A  I  I  N  N  S  M  C  N  H  L  F  L  K  Y  S
    ——————————— Protease F.CDS ———————————

    TTTCCGCAAGGACATCTTTGGAGACATGGTTTGTGCTGGCAATGCCCAAG
651 ---------+---------+---------+---------+---------+ 700
    AAAGGCGTTCCTGTAGAAACCTCTGTACCAAACACGACCGTTACGGGTTC
      F  R  K  D  I  F  G  D  M  V  C  A  G  N  A  Q
    ——————————— Protease F.CDS ———————————
```

FIG. 13(C)

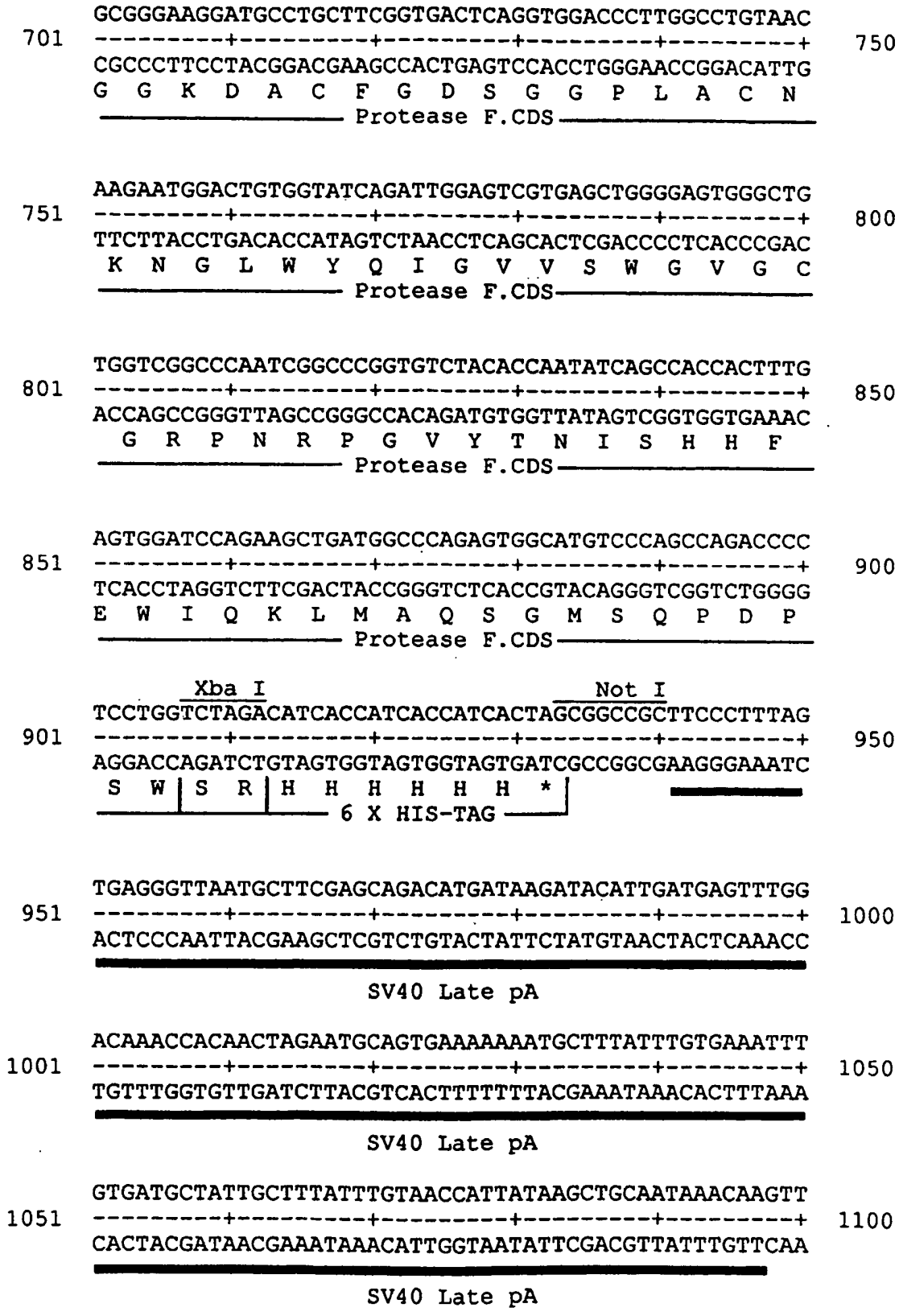

GCGGGAAGGATGCCTGCTTCGGTGACTCAGGTGGACCCTTGGCCTGTAAC
701
750
CGCCCTTCCTACGGACGAAGCCACTGAGTCCACCTGGGAACCGGACATTG
G G K D A C F G D S G G P L A C N
Protease F.CDS
AAGAATGGACTGTGGTATCAGATTGGAGTCGTGAGCTGGGGAGTGGGCTG
751
800
TTCTTACCTGACACCATAGTCTAACCTCAGCACTCGACCCCTCACCCGAC
K N G L W Y Q I G V V S W G V G C
Protease F.CDS
TGGTCGGCCCAATCGGCCCGGTGTCTACACCAATATCAGCCACCACTTTG
801
850
ACCAGCCGGGTTAGCCGGGCCACAGATGTGGTTATAGTCGGTGGTGAAAC
G R P N R P G V Y T N I S H H F
Protease F.CDS
AGTGGATCCAGAAGCTGATGGCCCAGAGTGGCATGTCCCAGCCAGACCCC
851
900
TCACCTAGGTCTTCGACTACCGGGTCTCACCGTACAGGGTCGGTCTGGGG
E W I Q K L M A Q S G M S Q P D P
Protease F.CDS
Xba I
Not I
TCCTGGTCTAGACATCACCATCACCATCACTAGCGGCCGCTTCCCTTTAG
901
950
AGGACCAGATCTGTAGTGGTAGTGGTAGTGATCGCCGGCGAAGGGAAATC
S W S R H H H H H H *
6 X HIS-TAG
TGAGGGTTAATGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGG
951
1000
ACTCCCAATTACGAAGCTCGTCTGTACTATTCTATGTAACTACTCAAACC
SV40 Late pA
ACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTT
1001
1050
TGTTTGGTGTTGATCTTACGTCACTTTTTTTACGAAATAAACACTTTAAA
SV40 Late pA
GTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTT
1051
1100
CACTACGATAACGAAATAAACATTGGTAATATTCGACGTTATTTGTTCAA
SV40 Late pA

FIG. 13(D)

```
        GAC
1101    ---   1103
        CTG
```

SEQ.ID.NO.:54

FIG. 14(A)

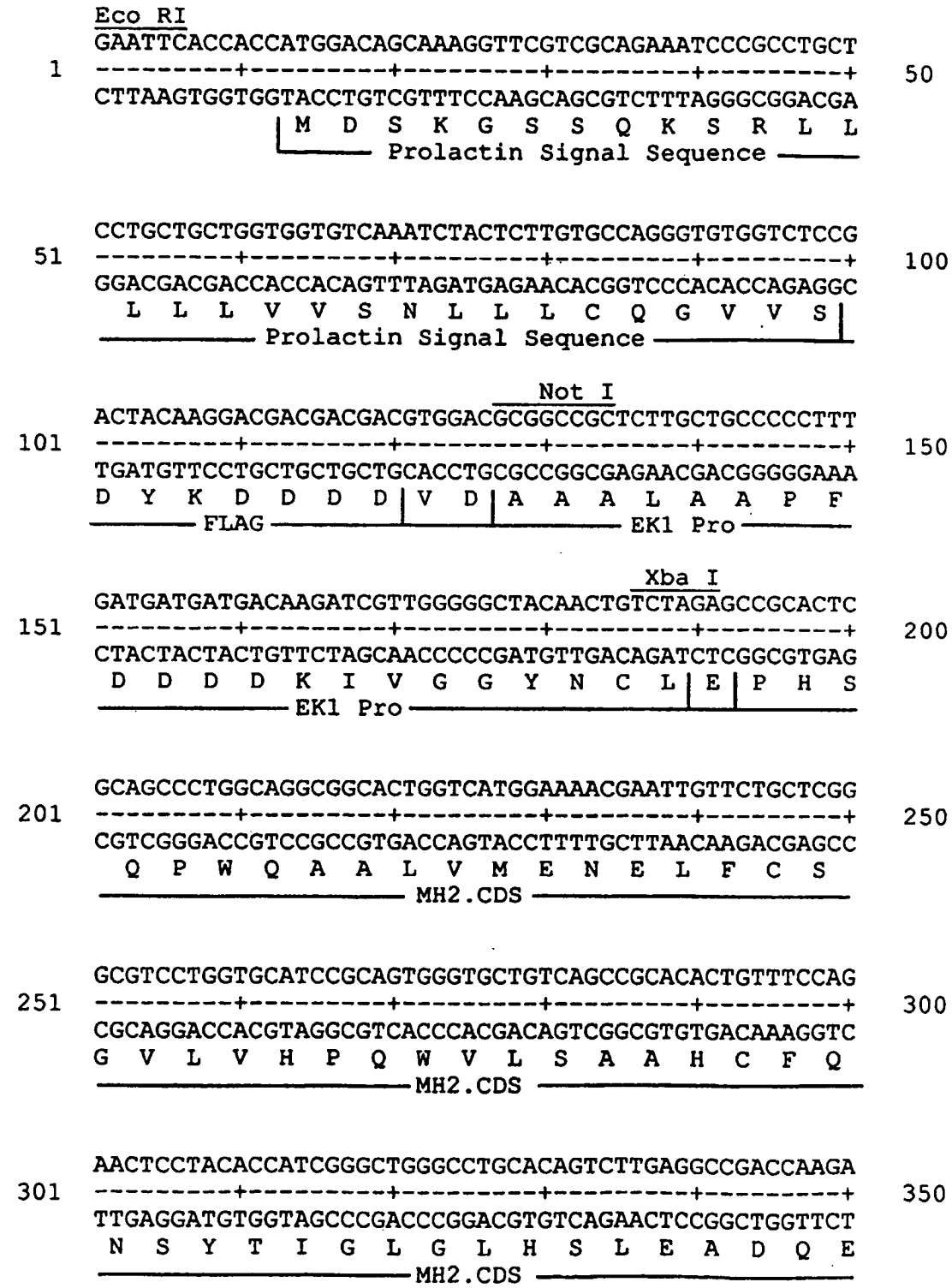
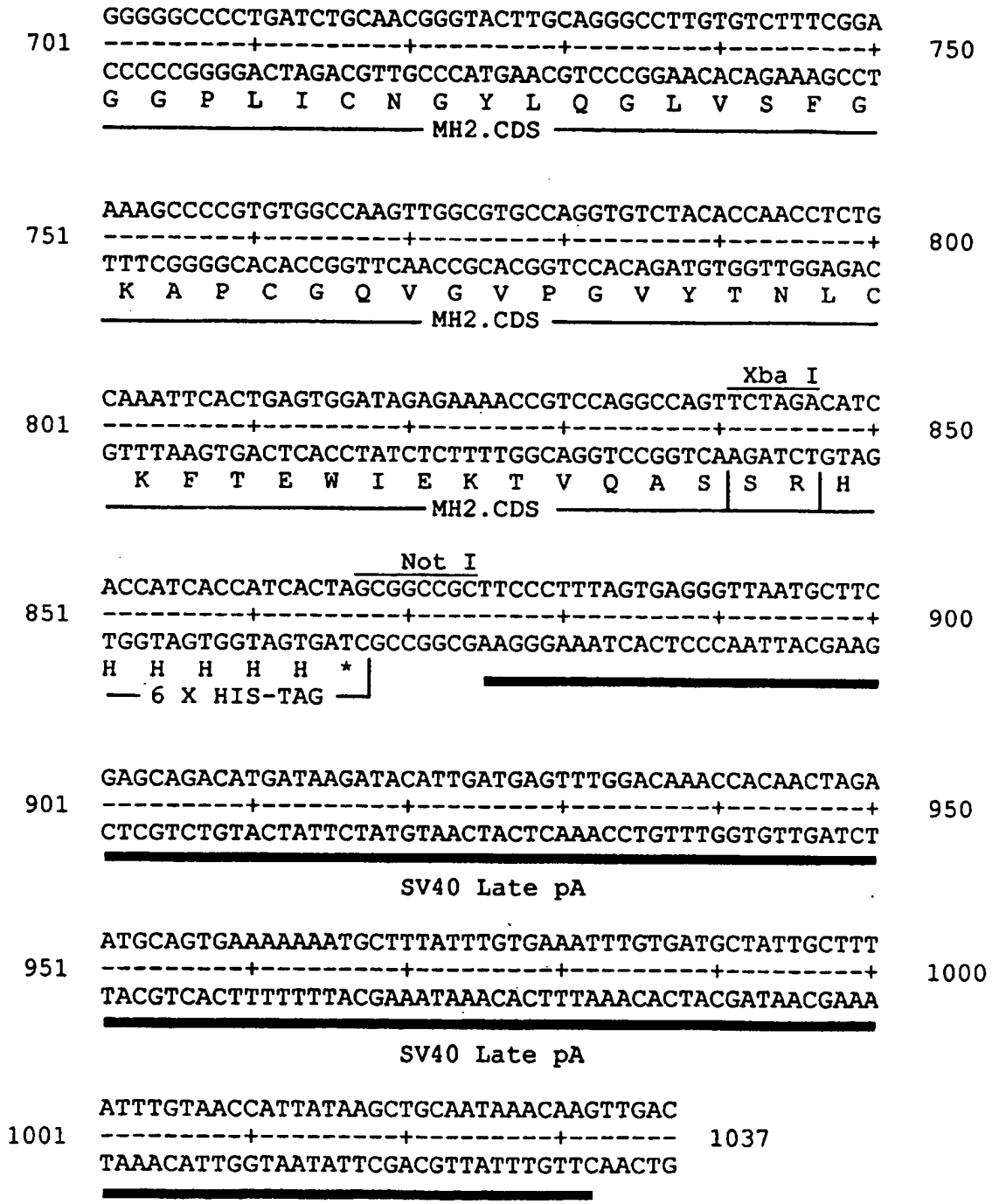


FIG. 14(B)

```
    GCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGAGT
351 ---------+---------+---------+---------+---------+ 400
    CGGTCCCTCGGTCTACCACCTCCGGTCGGAGAGGCATGCCGTGGGTCTCA
      P  G  S  Q  M  V  E  A  S  L  S  V  R  H  P  E
    ------------------------ MH2.CDS ------------------------

    ACAACAGACCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAA
401 ---------+---------+---------+---------+---------+ 450
    TGTTGTCTGGGAACGAGCGATTGCTGGAGTACGAGTAGTTCAACCTGCTT
    Y  N  R  P  L  L  A  N  D  L  M  L  I  K  L  D  E
    ------------------------ MH2.CDS ------------------------

    TCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTG
451 ---------+---------+---------+---------+---------+ 500
    AGGCACAGGCTCAGACTGTGGTAGGCCTCGTAGTCGTAACGAAGCGTCAC
     S  V  S  E  S  D  T  I  R  S  I  S  I  A  S  Q  C
    ------------------------ MH2.CDS ------------------------

    CCCTACCGCGGGGAACTCTTGCCTCGTTTCTGGCTGGGGTCTGCTGGCGA
501 ---------+---------+---------+---------+---------+ 550
    GGGATGGCGCCCCTTGAGAACGGAGCAAAGACCGACCCCAGACGACCGCT
      P  T  A  G  N  S  C  L  V  S  G  W  G  L  L  A
    ------------------------ MH2.CDS ------------------------

    ACGGCAGAATGCCTACCGTGCTGCAGTGCGTGAACGTGTCGGTGGTGTCT
551 ---------+---------+---------+---------+---------+ 600
    TGCCGTCTTACGGATGGCACGACGTCACGCACTTGCACAGCCACCACAGA
    N  G  R  M  P  T  V  L  Q  C  V  N  V  S  V  V  S
    ------------------------ MH2.CDS ------------------------

    GAGGAGGTCTGCAGTAAGCTCTATGACCCGCTGTACCACCCCAGCATGTT
601 ---------+---------+---------+---------+---------+ 650
    CTCCTCCAGACGTCATTCGAGATACTGGGCGACATGGTGGGGTCGTACAA
     E  E  V  C  S  K  L  Y  D  P  L  Y  H  P  S  M  F
    ------------------------ MH2.CDS ------------------------

    CTGCGCCGGCGGAGGGCACGACCAGAAGGACTCCTGCAACGGTGACTCTG
651 ---------+---------+---------+---------+---------+ 700
    GACGCGGCCGCCTCCCGTGCTGGTCTTCCTGAGGACGTTGCCACTGAGAC
      C  A  G  G  G  H  D  Q  K  D  S  C  N  G  D  S
    ------------------------ MH2.CDS ------------------------
```

FIG. 14(c)


GGGGGCCCCTGATCTGCAACGGGTACTTGCAGGGCCTTGTGTCTTTCGGA
701
750
CCCCCGGGGACTAGACGTTGCCCATGAACGTCCCGGAACACAGAAAGCCT
G G P L I C N G Y L Q G L V S F G
MH2.CDS
AAAGCCCCGTGTGGCCAAGTTGGCGTGCCAGGTGTCTACACCAACCTCTG
751
800
TTTCGGGGCACACCGGTTCAACCGCACGGTCCACAGATGTGGTTGGAGAC
K A P C G Q V G V P G V Y T N L C
MH2.CDS
Xba I
CAAATTCACTGAGTGGATAGAGAAAACCGTCCAGGCCAGTTCTAGACATC
801
850
GTTTAAGTGACTCACCTATCTCTTTTGGCAGGTCCGGTCAAGATCTGTAG
K F T E W I E K T V Q A S S R H
MH2.CDS
Not I
ACCATCACCATCACTAGCGGCCGCTTCCCTTTAGTGAGGGTTAATGCTTC
851
900
TGGTAGTGGTAGTGATCGCCGGCGAAGGGAAATCACTCCCAATTACGAAG
H H H H H *
6 X HIS-TAG
GAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGA
901
950
CTCGTCTGTACTATTCTATGTAACTACTCAAACCTGTTTGGTGTTGATCT
SV40 Late pA
ATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTT
951
1000
TACGTCACTTTTTTTACGAAATAAACACTTTAAACACTACGATAACGAAA
SV40 Late pA
ATTTGTAACCATTATAAGCTGCAATAAACAAGTTGAC
1001
1037
TAAACATTGGTAATATTCGACGTTATTTGTTCAACTG

SEQUENCE LISTING

<110> DARROW, ANDREW

QI, JENSON

ANDRADE-GORDON, PATRICIA

<120> ZYMOGEN ACTIVATION SYSTEM

<130> ORT-1028

<140>

<141>

<160> 60

<170> PATENTIN VER. 2.0

<210> 1

<211> 361

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE VECTORS.

<400> 1

GAATTCACCA CCATGGACAG CAAAGGTTCG TCGCAGAAAT CCCGCCTGCT CCTGCTGCTG 60

GTGGTGTCAA ATCTACTCTT GTGCCAGGGT GTGGTCTCCG ACTACAAGGA CGACGACGAC 120

GTGGACGCGG CCGCTCTTGC TGCCCCCTTT GATGATGATG ACAAGATCGT TGGGGGCTAT 180

GCTCTAGATA GCGGCCGCTT CCCTTTAGTG AGGGTTAATG CTTCGAGCAG ACATGATAAG 240

ATACATTGAT GAGTTTGGAC AAACCACAAC TAGAATGCAG TGAAAAAAAT GCTTTATTTG 300

TGAAATTTGT GATGCTATTG CTTTATTTGT AACCATTATA AGCTGCAATA AACAAGTTGA 360

C 361

<210> 2

<211> 301

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE VECTORS.

<400> 2

GAATTCACCA TGAATCCACT CCTGATCCTT ACCTTTGTGG CGGCCGCTCT TGCTGCCCCC 60

TTTGATGATG ATGACAAGAT CGTTGGGGGC TATTGTCTAG ATACCCCTAC GATGTGCCCG 120

ATTACGCCTA GCGGCCGCTT CCCTTTAGTG AGGGTTAATG CTTCGAGCAG ACATGATAAG 180

ATACATTGAT GAGTTTGGAC AAACCACAAC TAGAATGCAG TGAAAAAAAT GCTTTATTTG 240

TGAAATTTGT GATGCTATTG CTTTATTTGT AACCATTATA AGCTGCAATA AACAAGTTGA 300

C 301

<210> 3

<211> 484

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE VECTORS.

<400> 3

GAATTCACCA CCATGGACAG CAAAGGTTCG TCGCAGAAAT CCCGCCTGCT CCTGCTGCTG 60

GTGGTGTCAA ATCTACTCTT GTGCCAGGGT GTGGTCTCCG ACTACAAGGA CGACGACGAC 120

GTGGACGCGG CCGCTCTTGC TGCCCCCTTT ATCGAGGGGC GCATTGTGGA GGGCTCGGAT 180

CTAGATACCC CTACGATGTG CCCGATTACG CCGCTAGATA CCCCTACGAT GTGCCCGATT 240

ACGCCGCTAG ATACCACTAC GATGTGCCCG ATTACGCCGC TAGATACCCC TACGATGTGC 300

CCGATTACGC CTAGCGGCCG CTTCCCTTTA GTGAGGGTTA ATGCTTCGAG CAGACATGAT 360

AAGATACATT GATGAGTTTG GACAAACCAC AACTAGAATG CAGTGAAAAA AATGCTTTAT 420

TTGTGAAATT TGTGATGCTA TTGCTTTATT TGTAACCATT ATAAGCTGCA ATAAACAAGT 480

TGAC 484

<210> 4

<211> 382

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE

VECTORS.

<400> 4

GAATTCACCA CCATGGACAG CAAAGGTTCG TCGCAGAAAT CCCGCCTGCT CCTGCTGCTG 60

GTGGTGTCAA ATCTACTCTT GTGCCAGGGT GTGGTCTCCG ACTACAAGGA CGACGACGAC 120

GTGGACGCGG CCGCTCTTGC TGCCCCCTTT GATGATGATG ACAAGATCGT TGGGGGCTAC 180

MET ASP SER LYS GLY SER SER GLN LYS SER ARG LEU LEU LEU LEU LEU

1 5 10 15

VAL VAL SER ASN LEU LEU LEU CYS GLN GLY VAL VAL SER ASP TYR LYS

20 25 30

ASP ASP ASP ASP VAL ASP ALA ALA ALA LEU ALA ALA PRO PHE ASP ASP

35 40 45

ASP ASP LYS ILE VAL GLY GLY TYR ALA LEU GLU ALA GLY GLN TRP PRO

50 55 60

TRP GLN VAL SER ILE THR TYR GLU GLY VAL HIS VAL CYS GLY GLY SER

65 70 75 80

LEU VAL SER GLU GLN TRP VAL LEU SER ALA ALA HIS CYS PHE PRO SER

85 90 95

AACTGTCTAG ACATCACCAT CACCATCACT AGCGGCCGCT TCCCTTTAGT GAGGGTTAAT 240

GCTTCGAGCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA CTAGAATGCA 300

GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT 360

AAGCTGCAAT AAACAAGTTG AC 382

<210> 5

<211> 352

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE

VECTORS.

<400> 5

GAATTCACCA CCATGGCTTT CCTCTGGCTC CTCTCCTGCT GGGCCCTCCT GGGTACCACC 60

TTCGGCTGCG GGGTCCCCGA CTACAAGGAC GACGACGACG CGGCCGCTCT TGCTGCCCCC 120

TTTGATGATG ATGACAAGAT CGTTGGGGGC TATGCTCTAG ACATCACCAT CACCATCACT 180

AGCGGCCGCT TCCCTTTAGT GAGGGTTAAT GCTTCGAGCA GACATGATAA GATACATTGA 240

TGAGTTTGGA CAAACCACAA CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG 300

TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTG AC 352

<210> 6

<211> 385

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE

VECTORS.

<400> 6

GAATTCACCA CCATGGCTTT CCTCTGGCTC CTCTCCTGCT GGGCCCTCCT GGGTACCACC 60

TTCGGCTGCG GGGTCCCCGA CTACAAGGAC GACGACGACG CGGCCGCTCT TGCTGCCCCC 120

TTTGATGATG ATGACAAGAT CGTTGGGGGC TATGCTCTAG ATACCCCTAC GATGTGCCCG 180

ATTACGCCGC TAGACATCAC CATCACCATC ACTAGCGGCC GCTTCCCTTT AGTGAGGGTT 240

AATGCTTCGA GCAGACATGA TAAGATACAT TGATGAGTTT GGACAAACCA CAACTAGAAT 300

GCAGTGAAAA AAATGCTTTA TTTGTGAAAT TTGTGATGCT ATTGCTTTAT TTGTAACCAT 360

TATAAGCTGC AATAAACAAG TTGAC 385

<210> 7

<211> 1169

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE

WITH HOMO SAPIEN SERINE PROTEASE CATALYTIC DOMAIN

<400> 7

GAATTCACCA CCATGGACAG CAAAGGTTCG TCGCAGAAAT CCCGCCTGCT CCTGCTGCTG 60

GTGGTGTCAA ATCTACTCTT GTGCCAGGGT GTGGTCTCCG ACTACAAGGA CGACGACGAC 120

GTGGACGCGG CCGCTCTTGC TGCCCCCTTT GATGATGATG ACAAGATCGT TGGGGGCTAT 180

GCTCTAGAGG CCGGTCAGTG GCCCTGGCAG GTCAGCATCA CCTATGAAGG CGTCCATGTG 240

TGTGGTGGCT CTCTCGTGTC TGAGCAGTGG GTGCTGTCAG CTGCTCACTG CTTCCCCAGC 300

GAGCACCACA AGGAAGCCTA TGAGGTCAAG CTGGGGGCCC ACCAGCTAGA CTCCTACTCC 360

GAGGACGCCA AGGTCAGCAC CCTGAAGGAC ATCATCCCCC ACCCCAGCTA CCTCCAGGAG 420

GGCTCCCAGG GCGACATTGC ACTCCTCCAA CTCAGCAGAC CCATCACCTT CTCCCGCTAC 480

ATCCGGCCCA TCTGCCTCCC TGCAGCCAAC GCCTCCTTCC CCAACGGCCT CCACTGCACT 540

GTCACTGGCT GGGGTCATGT GGCCCCCTCA GTGAGCCTCC TGACGCCCAA GCCACTGCAG 600

CAACTCGAGG TGCCTCTGAT CAGTCGTGAG ACGTGTAACT GCCTGTACAA CATCGACGCC 660

AAGCCTGAGG AGCCGCACTT TGTCCAAGAG GACATGGTGT GTGCTGGCTA TGTGGAGGGG 720

GGCAAGGACG CCTGCCAGGG TGACTCTGGG GGCCCACTCT CCTGCCCTGT GGAGGGTCTC 780

TGGTACCTGA CGGGCATTGT GAGCTGGGGA GATGCCTGTG GGGCCCGCAA CAGGCCTGGT 840

GTGTACACTC TGGCCTCCAG CTATGCCTCC TGGATCCAAA GCAAGGTGAC AGAACTCCAG 900

CCTCGTGTGG TGCCCCAAAC CCAGGAGTCC CAGCCCGACA GCAACCTCTG TGGCAGCCAC 960

CTGGCCTTCA GCTCTAGACA TCACCATCAC CATCACTAGC GGCCGCTTCC CTTTAGTGAG 1020

GGTTAATGCT TCGAGCAGAC ATGATAAGAT ACATTGATGA GTTTGGACAA ACCACAACTA 1080

GAATGCAGTG AAAAAAATGC TTTATTTGTG AAATTTGTGA TGCTATTGCT TTATTTGTAA 1140

CCATTATAAG CTGCAATAAA CAAGTTGAC 1169

<210> 8

<211> 1142

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE

WITH HOMO SAPIEN SERINE PROTEASE CATALYTIC DOMAIN

<400> 8

GAATTCACCA CCATGGCTTT CCTCTGGCTC CTCTCCTGCT GGGCCCTCCT GGGTACCACC 60

TTCGGCTGCG GGGTCCCCGA CTACAAGGAC GACGACGACG CGGCCGCTCT TGCTGCCCCC 120

TTTGATGATG ATGACAAGAT CGTTGGGGGC TATGCTCTAG AGGCCGGTCA GTGGCCCTGG 180

CAGGTCAGCA TCACCTATGA AGGCGTCCAT GTGTGTGGTG GCTCTCTCGT GTCTGAGCAG 240

TGGGTGCTGT CAGCTGCTCA CTGCTTCCCC AGCGAGCACC ACAAGGAAGC CTATGAGGTC 300

AAGCTGGGGG CCCACCAGCT AGACTCCTAC TCCGAGGACG CCAAGGTCAG CACCCTGAAG 360

GACATCATCC CCCACCCCAG CTACCTCCAG GAGGGCTCCC AGGGCGACAT TGCACTCCTC 420

CAACTCAGCA GACCCATCAC CTTCTCCCGC TACATCCGGC CCATCTGCCT CCCTGCAGCC 480

AACGCCTCCT TCCCCAACGG CCTCCACTGC ACTGTCACTG GCTGGGGTCA TGTGGCCCCC 540

TCAGTGAGCC TCCTGACGCC CAAGCCACTG CAGCAACTCG AGGTGCCTCT GATCAGTCGT 600

GAGACGTGTA ACTGCCTGTA CAACATCGAC GCCAAGCCTG AGGAGCCGCA CTTTGTCCAA 660

GAGGACATGG TGTGTGCTGG CTATGTGGAG GGGGGCAAGG ACGCCTGCCA GGGTGACTCT 720

GGGGGCCCAC TCTCCTGCCC TGTGGAGGGT CTCTGGTACC TGACGGGCAT TGTGAGCTGG 780

GGAGATGCCT GTGGGGCCCG CAACAGGCCT GGTGTGTACA CTCTGGCCTC CAGCTATGCC 840

TCCTGGATCC AAAGCAAGGT GACAGAACTC CAGCCTCGTG TGGTGCCCCA AACCCAGGAG 900

TCCCAGCCCG ACAGCAACCT CTGTGGCAGC CACCTGGCCT TCAGCTCTAG ACATCACCAT 960

CACCATCACT AGCGGCCGCT TCCCTTTAGT GAGGGTTAAT GCTTCGAGCA GACATGATAA 1020

GATACATTGA TGAGTTTGGA CAAACCACAA CTAGAATGCA GTGAAAAAAA TGCTTTATTT 1080

GTGAAATTTG TGATGCTATT GCTTTATTTG TAACCATTAT AAGCTGCAAT AAACAAGTTG 1140

AC 1142

<210> 9

<211> 1049

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE
WITH HOMO SAPIEN SERINE PROTEASE CATALYTIC DOMAIN

<400> 9

GAATTCACCA CCATGGACAG CAAAGGTTCG TCGCAGAAAT CCCGCCTGCT CCTGCTGCTG 60

GTGGTGTCAA ATCTACTCTT GTGCCAGGGT GTGGTCTCCG ACTACAAGGA CGACGACGAC 120

GTGGACGCGG CCGCTCTTGC TGCCCCCTTT GATGATGATG ACAAGATCGT TGGGGGCTAC 180

AACTGTCTAG AACCCCATTC GCAGCCTTGG CAGGCGGCCT TGTTCCAGGG CCAGCAACTA 240

CTCTGTGGCG GTGTCCTTGT AGGTGGCAAC TGGGTCCTTA CAGCTGCCCA CTGTAAAAAA 300

CCGAAATACA CAGTACGCCT GGGAGACCAC AGCCTACAGA ATAAAGATGG CCCAGAGCAA 360

GAAATACCTG TGGTTCAGTC CATCCCACAC CCCTGCTACA ACAGCAGCGA TGTGGAGGAC 420

CACAACCATG ATCTGATGCT TCTTCAACTG CGTGACCAGG CATCCCTGGG GTCCAAAGTG 480

AAGCCCATCA GCCTGGCAGA TCATTGCACC CAGCCTGGCC AGAAGTGCAC CGTCTCAGGC 540

TGGGGCACTG TCACCAGTCC CCGAGAGAAT TTTCCTGACA CTCTCAACTG TGCAGAAGTA 600

AAAATCTTTC CCCAGAAGAA GTGTGAGGAT GCTTACCCGG GGCAGATCAC AGATGGCATG 660

GTCTGTGCAG GCAGCAGCAA AGGGGCTGAC ACGTGCCAGG GCGATTCTGG AGGCCCCCTG 720

GTGTGTGATG GTGCACTCCA GGGCATCACA TCCTGGGGCT CAGACCCCTG TGGGAGGTCC 780

GACAAACCTG GCGTCTATAC CAACATCTGC CGCTACCTGG ACTGGATCAA GAAGATCATA 840

GGCAGCAAGG GCTCTAGACA TCACCATCAC CATCACTAGC GGCCGCTTCC CTTTAGTGAG 900

GGTTAATGCT TCGAGCAGAC ATGATAAGAT ACATTGATGA GTTTGGACAA ACCACAACTA 960

GAATGCAGTG AAAAAAATGC TTTATTTGTG AAATTTGTGA TGCTATTGCT TTATTTGTAA 1020

CCATTATAAG CTGCAATAAA CAAGTTGAC 1049

<210> 10

<211> 1052

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE

WITH HOMO SAPIEN SERINE PROTEASE CATALYTIC DOMAIN

<400> 10

GAATTCACCA CCATGGACAG CAAAGGTTCG TCGCAGAAAT CCCGCCTGCT CCTGCTGCTG 60

GTGGTGTCAA ATCTACTCTT GTGCCAGGGT GTGGTCTCCG ACTACAAGGA CGACGACGAC 120

GTGGACGCGG CCGCTCTTGC TGCCCCCTTT GATGATGATG ACAAGATCGT TGGGGGCTAC 180

AACTGTCTAG AAAAGCACTC CCAGCCCTGG CAGGCAGCCC TGTTCGAGAA GACGCGGCTA 240

CTCTGTGGGG CGACGCTCAT CGCCCCCAGA TGGCTCCTGA CAGCAGCCCA CTGCCTCAAG 300

CCCCGCTACA TAGTTCACCT GGGGCAGCAC AACCTCCAGA AGGAGGAGGG CTGTGAGCAG 360

ACCCGGACAG CCACTGAGTC CTTCCCCCAC CCCGGCTTCA ACAACAGCCT CCCCAACAAA 420

GACCACCGCA ATGACATCAT GCTGGTGAAG ATGGCATCGC CAGTCTCCAT CACCTGGGCT 480

GTGCGACCCC TCACCCTCTC CTCACGCTGT GTCACTGCTG GCACCAGCTG CCTCATTTCC 540

GGCTGGGGCA GCACGTCCAG CCCCCAGTTA CGCCTGCCTC ACACCTTGCG ATGCGCCAAC 600

ATCACCATCA TTGAGCACCA GAAGTGTGAG AACGCCTACC CCGGCAACAT CACAGACACC 660

ATGGTGTGTG CCAGCGTGCA GGAAGGGGGC AAGGACTCCT GCCAGGGTGA CTCCGGGGGC 720

CCTCTGGTCT GTAACCAGTC TCTTCAAGGC ATTATCTCCT GGGGCCAGGA TCCGTGTGCG 780

ATCACCCGAA AGCCTGGTGT CTACACGAAA GTCTGCAAAT ATGTGGACTG GATCCAGGAG 840

ACGATGAAGA ACAATTCTAG ACATCACCAT CACCATCACT AGCGGCCGCT TCCCTTTAGT 900

GAGGGTTAAT GCTTCGAGCA GACATGATAA GATACATTGA TGAGTTTGGA CAAACCACAA 960

CTAGAATGCA GTGAAAAAAA TGCTTTATTT GTGAAATTTG TGATGCTATT GCTTTATTTG 1020

TAACCATTAT AAGCTGCAAT AAACAAGTTG AC 1052

<210> 11

<211> 328

<212> PRT

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE

WITH HOMO SAPIEN SERINE PROTEASE CATALYTIC DOMAIN

<400> 11

GLU HIS HIS LYS GLU ALA TYR GLU VAL LYS LEU GLY ALA HIS GLN LEU

100 105 110

ASP SER TYR SER GLU ASP ALA LYS VAL SER THR LEU LYS ASP ILE ILE

115 120 125

PRO HIS PRO SER TYR LEU GLN GLU GLY SER GLN GLY ASP ILE ALA LEU

130 135 140

LEU GLN LEU SER ARG PRO ILE THR PHE SER ARG TYR ILE ARG PRO ILE

145 150 155 160

CYS LEU PRO ALA ALA ASN ALA SER PHE PRO ASN GLY LEU HIS CYS THR

165 170 175

VAL THR GLY TRP GLY HIS VAL ALA PRO SER VAL SER LEU LEU THR PRO

180 185 190

LYS PRO LEU GLN GLN LEU GLU VAL PRO LEU ILE SER ARG GLU THR CYS

195 200 205

ASN CYS LEU TYR ASN ILE ASP ALA LYS PRO GLU GLU PRO HIS PHE VAL

210 215 220

GLN GLU ASP MET VAL CYS ALA GLY TYR VAL GLU GLY GLY LYS ASP ALA

225 230 235 240

CYS GLN GLY ASP SER GLY GLY PRO LEU SER CYS PRO VAL GLU GLY LEU

245 250 255

TRP TYR LEU THR GLY ILE VAL SER TRP GLY ASP ALA CYS GLY ALA ARG

260 265 270

ASN ARG PRO GLY VAL TYR THR LEU ALA SER SER TYR ALA SER TRP ILE

275 280 285

GLN SER LYS VAL THR GLU LEU GLN PRO ARG VAL VAL PRO GLN THR GLN

290 295 300

GLU SER GLN PRO ASP SER ASN LEU CYS GLY SER HIS LEU ALA PHE SER

305 310 315 320

SER ARG HIS HIS HIS HIS HIS HIS

325

<210> 12

<211> 319

<212> PRT

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE

WITH HOMO SAPIEN SERINE PROTEASE CATALYTIC DOMAIN

<400> 12

MET ALA PHE LEU TRP LEU LEU SER CYS TRP ALA LEU LEU GLY THR THR

1 5 10 15

PHE GLY CYS GLY VAL PRO ASP TYR LYS ASP ASP ASP ASP ALA ALA ALA

20 25 30

LEU ALA ALA PRO PHE ASP ASP ASP ASP LYS ILE VAL GLY GLY TYR ALA

35 40 45

LEU GLU ALA GLY GLN TRP PRO TRP GLN VAL SER ILE THR TYR GLU GLY

50 55 60

VAL HIS VAL CYS GLY GLY SER LEU VAL SER GLU GLN TRP VAL LEU SER

65 70 75 80

ALA ALA HIS CYS PHE PRO SER GLU HIS HIS LYS GLU ALA TYR GLU VAL

85 90 95

LYS LEU GLY ALA HIS GLN LEU ASP SER TYR SER GLU ASP ALA LYS VAL

100 105 110

SER THR LEU LYS ASP ILE ILE PRO HIS PRO SER TYR LEU GLN GLU GLY

115 120 125

SER GLN GLY ASP ILE ALA LEU LEU GLN LEU SER ARG PRO ILE THR PHE

130 135 140

SER ARG TYR ILE ARG PRO ILE CYS LEU PRO ALA ALA ASN ALA SER PHE

145 150 155 160

PRO ASN GLY LEU HIS CYS THR VAL THR GLY TRP GLY HIS VAL ALA PRO

165 170 175

SER VAL SER LEU LEU THR PRO LYS PRO LEU GLN GLN LEU GLU VAL PRO

180 185 190

LEU ILE SER ARG GLU THR CYS ASN CYS LEU TYR ASN ILE ASP ALA LYS

195 200 205

PRO GLU GLU PRO HIS PHE VAL GLN GLU ASP MET VAL CYS ALA GLY TYR

210 215 220

VAL GLU GLY GLY LYS ASP ALA CYS GLN GLY ASP SER GLY GLY PRO LEU

225 230 235 240

SER CYS PRO VAL GLU GLY LEU TRP TYR LEU THR GLY ILE VAL SER TRP

245 250 255

GLY ASP ALA CYS GLY ALA ARG ASN ARG PRO GLY VAL TYR THR LEU ALA

260 265 270

SER SER TYR ALA SER TRP ILE GLN SER LYS VAL THR GLU LEU GLN PRO

275 280 285

ARG VAL VAL PRO GLN THR GLN GLU SER GLN PRO ASP SER ASN LEU CYS

290 295 300

GLY SER HIS LEU ALA PHE SER SER ARG HIS HIS HIS HIS HIS HIS

305 310 315

<210> 13

<211> 288

<212> PRT

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE
WITH HOMO SAPIEN SERINE PROTEASE CATALYTIC DOMAIN

<400> 13

MET ASP SER LYS GLY SER SER GLN LYS SER ARG LEU LEU LEU LEU LEU
1 5 10 15

VAL VAL SER ASN LEU LEU LEU CYS GLN GLY VAL VAL SER ASP TYR LYS
20 25 30

ASP ASP ASP ASP VAL ASP ALA ALA ALA LEU ALA ALA PRO PHE ASP ASP
35 40 45

ASP ASP LYS ILE VAL GLY GLY TYR ASN CYS LEU GLU PRO HIS SER GLN

50 55 60

PRO TRP GLN ALA ALA LEU PHE GLN GLY GLN GLN LEU LEU CYS GLY GLY

65 70 75 80

VAL LEU VAL GLY GLY ASN TRP VAL LEU THR ALA ALA HIS CYS LYS LYS

85 90 95

PRO LYS TYR THR VAL ARG LEU GLY ASP HIS SER LEU GLN ASN LYS ASP

100 105 110

GLY PRO GLU GLN GLU ILE PRO VAL VAL GLN SER ILE PRO HIS PRO CYS

115 120 125

TYR ASN SER SER ASP VAL GLU ASP HIS ASN HIS ASP LEU MET LEU LEU

130 135 140

GLN LEU ARG ASP GLN ALA SER LEU GLY SER LYS VAL LYS PRO ILE SER

145 150 155 160

LEU ALA ASP HIS CYS THR GLN PRO GLY GLN LYS CYS THR VAL SER GLY

165 170 175

TRP GLY THR VAL THR SER PRO ARG GLU ASN PHE PRO ASP THR LEU ASN

180 185 190

CYS ALA GLU VAL LYS ILE PHE PRO GLN LYS LYS CYS GLU ASP ALA TYR

195 200 205

PRO GLY GLN ILE THR ASP GLY MET VAL CYS ALA GLY SER SER LYS GLY

210 215 220

ALA ASP THR CYS GLN GLY ASP SER GLY GLY PRO LEU VAL CYS ASP GLY

225 230 235 240

ALA LEU GLN GLY ILE THR SER TRP GLY SER ASP PRO CYS GLY ARG SER

245 250 255

ASP LYS PRO GLY VAL TYR THR ASN ILE CYS ARG TYR LEU ASP TRP ILE

260 265 270

LYS LYS ILE ILE GLY SER LYS GLY SER ARG HIS HIS HIS HIS HIS HIS

275 280 285

<210> 14

<211> 289

<212> PRT

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE
WITH HOMO SAPIEN SERINE PROTEASE CATALYTIC DOMAIN

<400> 14

MET ASP SER LYS GLY SER SER GLN LYS SER ARG LEU LEU LEU LEU LEU
1 5 10 15

VAL VAL SER ASN LEU LEU LEU CYS GLN GLY VAL VAL SER ASP TYR LYS
20 25 30

ASP ASP ASP ASP VAL ASP ALA ALA ALA LEU ALA ALA PRO PHE ASP ASP
35 40 45

ASP ASP LYS ILE VAL GLY GLY TYR ASN CYS LEU GLU LYS HIS SER GLN

50 55 60

PRO TRP GLN ALA ALA LEU PHE GLU LYS THR ARG LEU LEU CYS GLY ALA

65 70 75 80

THR LEU ILE ALA PRO ARG TRP LEU LEU THR ALA ALA HIS CYS LEU LYS

85 90 95

PRO ARG TYR ILE VAL HIS LEU GLY GLN HIS ASN LEU GLN LYS GLU GLU

100 105 110

GLY CYS GLU GLN THR ARG THR ALA THR GLU SER PHE PRO HIS PRO GLY

115 120 125

PHE ASN ASN SER LEU PRO ASN LYS ASP HIS ARG ASN ASP ILE MET LEU

130 135 140

VAL LYS MET ALA SER PRO VAL SER ILE THR TRP ALA VAL ARG PRO LEU

145 150 155 160

THR LEU SER SER ARG CYS VAL THR ALA GLY THR SER CYS LEU ILE SER

165 170 175

GLY TRP GLY SER THR SER SER PRO GLN LEU ARG LEU PRO HIS THR LEU

180 185 190

ARG CYS ALA ASN ILE THR ILE ILE GLU HIS GLN LYS CYS GLU ASN ALA

195 200 205

TYR PRO GLY ASN ILE THR ASP THR MET VAL CYS ALA SER VAL GLN GLU

210 215 220

GLY GLY LYS ASP SER CYS GLN GLY ASP SER GLY GLY PRO LEU VAL CYS

225 230 235 240

ASN GLN SER LEU GLN GLY ILE ILE SER TRP GLY GLN ASP PRO CYS ALA

245 250 255

ILE THR ARG LYS PRO GLY VAL TYR THR LYS VAL CYS LYS TYR VAL ASP

260 265 270

TRP ILE GLN GLU THR MET LYS ASN ASN SER ARG HIS HIS HIS HIS HIS

275 280 285

HIS

<210> 15

<211> 9

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 15

CTAGATAGC 9

<210> 16

<211> 9

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 16

GGCCGCTAT 9

<210> 17

<211> 36

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 17

CTAGATACCC CTACGATGTG CCCGATTACG CCTAGC 36

<210> 18

<211> 36

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 18

GGCCGCTAGG CGTAATCGGG CACATCGTAG GGGTAT 36

<210> 19

<211> 33

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 19

CTAGATACCC CTACGATGTG CCCGATTACG CCG 33

<210> 20

<211> 33

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 20

CTAGCGGCGT AATCGGGCAC ATCGTAGGGG TAT 33

<210> 21

<211> 27

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 21

CTAGACATCA CCATCACCAT CACTAGC 27

<210> 22

<211> 27

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 22

GGCCGCTAGT GATGGTGATG GTGATGT 27

<210> 23

<211> 34

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 23

TGAATTCACC ACCATGGACA GCAAAGGTTC GTCG 34

<210> 24

<211> 30

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 24

CAGAAAGGGT CCCGCCTGCT CCTGCTGCTG 30

<210> 25

<211> 30

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 25

GTGGTGTCAA ATCTACTCTT GTGCCAGGGT 30

<210> 26

<211> 30

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 26

GTGGTCTCCG ACTACAAGGA CGACGACGAC 30

<210> 27

<211> 21

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 27

GTGGACGCGG CCGCATTATT A 21

<210> 28

<211> 35

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 28

TAATAATGCG GCCGCGTCCA CGTCGTCGTC GTCCT 35

<210> 29

<211> 21

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 29

TGTAGTCGGA GACCACACCC T 21

<210> 30

<211> 30

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 30

GGCACAAGAG TAGATTTGAC ACCACCAGCA 30

<210> 31

<211> 30

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 31

GCAGGAGCAG GCGGGACCCT TTCTGCGACG 30

<210> 32

<211> 29

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 32

AACCTTTGCT GTCCATGGTG GTGAATTCA 29

<210> 33

<211> 40

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 33

AATTCACCAT GAATCCACTC CTGATCCTTA CCTTTGTGGC 40

<210> 34

<211> 40

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 34

GGCCGCCACA AAGGTAAGGA TCAGGAGTGG ATTCATGGTG 40

<210> 35

<211> 55

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 35

AATTCACCAC CATGGCTTTC CTCTGGCTCC TCTCCTGCTG GGCCCTCCTG GGTAC 55

<210> 36

<211> 47

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 36

CCAGGAGGGC CCAGCAGGAG AGGAGCCAGA GGAAAGCCAT GGTGGTG 47

<210> 37

<211> 45

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 37

CACCTTCGGC TGCGGGGTCC CCGACTACAA GGACGACGAC GACGC 45

<210> 38

<211> 53

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 38

GGCCGCGTCG TCGTCGTCCT TGTAGTCGGG GACCCCGCAG CCGAAGGTGG TAC 53

<210> 39

<211> 29

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 39

GTGGCGGCCG CTCTTGCTGC CCCCTTTGA 29

<210> 40

<211> 28

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 40

TTCTCTAGAC AGTTGTAGCC CCCAACGA 28

<210> 41

<211> 55

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 41

GGCCGCTCTT GCTGCCCCCT TTGATGATGA TGACAAGATC GTTGGGGGCT ATGCT 55

<210> 42

<211> 55

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 42

CTAGAGCATA GCCCCCAACG ATCTTGTCAT CATCATCAAA GGGGGCAGCA AGAGC 55

<210> 43

<211> 55

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 43

GGCCGCTCTT GCTGCCCCCT TTGATGATGA TGACAAGATC GTTGGGGGCT ATTGT 55

<210> 44

<211> 55

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 44

CTAGACAATA GCCCCCAACG ATCTTGTCAT CATCATCAAA GGGGGCAGCA AGAGC 55

<210> 45

<211> 52

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 45

GGCCGCTCTT GCTGCCCCCT TTATCGAGGG GCGCATTGTG GAGGGCTCGG AT 52

<210> 46

<211> 52

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 46

CTAGATCCGA GCCCTCCACA ATGCGCCCCT CGATAAAGGG GGCAGCAAGA GC 52

<210> 47

<211> 32

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 47

AGCAGTCTAG AGGCCGGTCA GTGGCCCTGG CA 32

<210> 48

<211> 28

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 48

GCTGGTCTAG AGCTGAAGGC CAGGTGGC 28

<210> 49

<211> 29

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 49

GGTATCTAGA GCCCTTGCTG CCTATGATC 29

<210> 50

<211> 30

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 50

ACTGTCTAGA ACCCCATTCG CAGCCTTGGC 30

<210> 51

<211> 32

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 51

TCGATCTAGA AAAGCACTCC CAGCCCTGGC AG 32

<210> 52

<211> 32

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE:

OLIGONUCLEOTIDE

<400> 52

GTCCTCTAGA ATTGTTCTTC ATCGTCTCCT GG 32

<210> 53

<211> 306

<212> PRT

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: FUSION GENE OF HUMAN PROTEASE F IN CFEK2 ZYMOGEN VECTOR

<400> 53

MET ALA PHE LEU TRP LEU LEU SER CYS TRP ALA LEU LEU GLY THR THR
1 5 10 15

PHE GLY CYS GLY VAL PRO ASP TYR LYS ASP ASP ASP ASP ALA ALA ALA
20 25 30

LEU ALA ALA PRO PHE ASP ASP ASP ASP LYS ILE VAL GLY GLY TYR ALA
35 40 45

LEU GLU LEU GLY ARG TRP PRO TRP GLN GLY SER LEU ARG LEU TRP ASP
50 55 60

SER HIS VAL CYS GLY VAL SER LEU LEU SER HIS ARG TRP ALA LEU THR

65 70 75 80

ALA ALA HIS CYS PHE GLU THR TYR SER ASP LEU SER ASP PRO SER GLY

85 90 95

TRP MET VAL GLN PHE GLY GLN LEU THR SER MET PRO SER PHE TRP SER

100 105 110

LEU GLN ALA TYR TYR ASN ARG TYR PHE VAL SER ASN ILE TYR LEU SER

115 120 125

PRO ARG TYR LEU GLY ASN SER PRO TYR ASP ILE ALA LEU VAL LYS LEU

130 135 140

SER ALA PRO VAL THR TYR THR LYS HIS ILE GLN PRO ILE CYS LEU GLN

145 150 155 160

ALA SER THR PHE GLU PHE GLU ASN ARG THR ASP CYS TRP VAL THR GLY

165 170 175

TRP GLY TYR ILE LYS GLU ASP GLU ALA LEU PRO SER PRO HIS THR LEU

180 185 190

GLN GLU VAL GLN VAL ALA ILE ILE ASN ASN SER MET CYS ASN HIS LEU

195 200 205

PHE LEU LYS TYR SER PHE ARG LYS ASP ILE PHE GLY ASP MET VAL CYS

210 215 220

ALA GLY ASN ALA GLN GLY GLY LYS ASP ALA CYS PHE GLY ASP SER GLY

225 230 235 240

GLY PRO LEU ALA CYS ASN LYS ASN GLY LEU TRP TYR GLN ILE GLY VAL

245 250 255

VAL SER TRP GLY VAL GLY CYS GLY ARG PRO ASN ARG PRO GLY VAL TYR

260 265 270

THR ASN ILE SER HIS HIS PHE GLU TRP ILE GLN LYS LEU MET ALA GLN

275 280 285

SER GLY MET SER GLN PRO ASP PRO SER TRP SER ARG HIS HIS HIS HIS

290 295 300

HIS HIS

305

<210> 54

<211> 284

<212> PRT

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: HUMAN MH2
PROTEASE IN PFEK ZYMOGEN VECTOR

<400> 54

MET ASP SER LYS GLY SER SER GLN LYS SER ARG LEU LEU LEU LEU LEU
1 5 10 15

VAL VAL SER ASN LEU LEU LEU CYS GLN GLY VAL VAL SER ASP TYR LYS
20 25 30

ASP ASP ASP ASP VAL ASP ALA ALA ALA LEU ALA ALA PRO PHE ASP ASP
35 40 45

ASP ASP LYS ILE VAL GLY GLY TYR ASN CYS LEU GLU PRO HIS SER GLN

50 55 60

PRO TRP GLN ALA ALA LEU VAL MET GLU ASN GLU LEU PHE CYS SER GLY

65 70 75 80

VAL LEU VAL HIS PRO GLN TRP VAL LEU SER ALA ALA HIS CYS PHE GLN

85 90 95

ASN SER TYR THR ILE GLY LEU GLY LEU HIS SER LEU GLU ALA ASP GLN

100 105 110

GLU PRO GLY SER GLN MET VAL GLU ALA SER LEU SER VAL ARG HIS PRO

115 120 125

GLU TYR ASN ARG PRO LEU LEU ALA ASN ASP LEU MET LEU ILE LYS LEU

130 135 140

ASP GLU SER VAL SER GLU SER ASP THR ILE ARG SER ILE SER ILE ALA

145 150 155 160

SER GLN CYS PRO THR ALA GLY ASN SER CYS LEU VAL SER GLY TRP GLY

165 170 175

LEU LEU ALA ASN GLY ARG MET PRO THR VAL LEU GLN CYS VAL ASN VAL

180 185 190

SER VAL VAL SER GLU GLU VAL CYS SER LYS LEU TYR ASP PRO LEU TYR

195 200 205

HIS PRO SER MET PHE CYS ALA GLY GLY GLY HIS ASP GLN LYS ASP SER

210 215 220

CYS ASN GLY ASP SER GLY GLY PRO LEU ILE CYS ASN GLY TYR LEU GLN

225 230 235 240

GLY LEU VAL SER PHE GLY LYS ALA PRO CYS GLY GLN VAL GLY VAL PRO

245 250 255

GLY VAL TYR THR ASN LEU CYS LYS PHE THR GLU TRP ILE GLU LYS THR

260 265 270

VAL GLN ALA SER SER ARG HIS HIS HIS HIS HIS HIS

275 280

<210> 55

<211> 30

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: PCR PRIMER

<400> 55

AGGATCTAGA GCCGCACTCG CAGCCCTGGC 30

<210> 56

<211> 30

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: PCR PRIMER

<400> 56

CCCATCTAGA ACTGGCCTGG ACGGTTTTCT 30

<210> 57

<211> 32

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: PCR PRIMER

<400> 57

AGGATCTAGA ACTCGGGCGT TGGCCGTGGC AG 32

<210> 58

<211> 30

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: PCR PRIMER

<400> 58

AGAGTCTAGA CCAGGAGGGG TCTGGCTGGG 30

<210> 59

<211> 1103

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: NUCLEIC ACID
SEQUENCE OF HUMAN PROTEASE F IN CFEK2 ZYMOGEN
VECTOR

<400> 59

GAATTCACCA CCATGGCTTT CCTCTGGCTC CTCTCCTGCT GGGCCCTCCT GGGTACCACC 60

TTCGGCTGCG GGGTCCCCGA CTACAAGGAC GACGACGACG CGGCCGCTCT TGCTGCCCCC 120

TTTGATGATG ATGACAAGAT CGTTGGGGGC TATGCTCTAG AACTCGGGCG TTGGCCGTGG 180

CAGGGGAGCC TGCGCCTGTG GGATTCCCAC GTATGCGGAG TGAGCCTGCT CAGCCACCGC 240

TGGGCACTCA CGGCGGCGCA CTGCTTTGAA ACCTATAGTG ACCTTAGTGA TCCCTCCGGG 300

TGGATGGTCC AGTTTGGCCA GCTGACTTCC ATGCCATCCT TCTGGAGCCT GCAGGCCTAC 360

TACAACCGTT ACTTCGTATC GAATATCTAT CTGAGCCCTC GCTACCTGGG GAATTCACCC 420

TATGACATTG CCTTGGTGAA GCTGTCTGCA CCTGTCACCT ACACTAAACA CATCCAGCCC 480

ATCTGTCTCC AGGCCTCCAC ATTTGAGTTT GAGAACCGGA CAGACTGCTG GGTGACTGGC 540

TGGGGGTACA TCAAAGAGGA TGAGGCACTG CCATCTCCCC ACACCCTCCA GGAAGTTCAG 600

GTCGCCATCA TAAACAACTC TATGTGCAAC CACCTCTTCC TCAAGTACAG TTTCCGCAAG 660

GACATCTTTG GAGACATGGT TTGTGCTGGC AATGCCCAAG GCGGGAAGGA TGCCTGCTTC 720

GGTGACTCAG GTGGACCCTT GGCCTGTAAC AAGAATGGAC TGTGGTATCA GATTGGAGTC 780

GTGAGCTGGG GAGTGGGCTG TGGTCGGCCC AATCGGCCCG GTGTCTACAC CAATATCAGC 840

CACCACTTTG AGTGGATCCA GAAGCTGATG GCCCAGAGTG GCATGTCCCA GCCAGACCCC 900

TCCTGGTCTA GACATCACCA TCACCATCAC TAGCGGCCGC TTCCCTTTAG TGAGGGTTAA 960

TGCTTCGAGC AGACATGATA AGATACATTG ATGAGTTTGG ACAAACCACA ACTAGAATGC 1020

AGTGAAAAAA ATGCTTTATT TGTGAAATTT GTGATGCTAT TGCTTTATTT GTAACCATTA 1080

TAAGCTGCAA TAAACAAGTT GAC 1103

<210> 60

<211> 1037

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> DESCRIPTION OF ARTIFICIAL SEQUENCE: NUCLEIC ACID
SEQUENCE OF HUMAN MH2 PROTEASE IN PFEK ZYMOGEN
VECTOR

<400> 60

GAATTCACCA CCATGGACAG CAAAGGTTCG TCGCAGAAAT CCCGCCTGCT CCTGCTGCTG 60

GTGGTGTCAA ATCTACTCTT GTGCCAGGGT GTGGTCTCCG ACTACAAGGA CGACGACGAC 120

GTGGACGCGG CCGCTCTTGC TGCCCCCTTT GATGATGATG ACAAGATCGT TGGGGGCTAC 180

AACTGTCTAG AGCCGCACTC GCAGCCCTGG CAGGCGGCAC TGGTCATGGA AAACGAATTG 240

TTCTGCTCGG GCGTCCTGGT GCATCCGCAG TGGGTGCTGT CAGCCGCACA CTGTTTCCAG 300

AACTCCTACA CCATCGGGCT GGGCCTGCAC AGTCTTGAGG CCGACCAAGA GCCAGGGAGC 360

CAGATGGTGG AGGCCAGCCT CTCCGTACGG CACCCAGAGT ACAACAGACC CTTGCTCGCT 420

AACGACCTCA TGCTCATCAA GTTGGACGAA TCCGTGTCCG AGTCTGACAC CATCCGGAGC 480

ATCAGCATTG CTTCGCAGTG CCCTACCGCG GGGAACTCTT GCCTCGTTTC TGGCTGGGGT 540

CTGCTGGCGA ACGGCAGAAT GCCTACCGTG CTGCAGTGCG TGAACGTGTC GGTGGTGTCT 600

GAGGAGGTCT GCAGTAAGCT CTATGACCCG CTGTACCACC CCAGCATGTT CTGCGCCGGC 660

GGAGGGCACG ACCAGAAGGA CTCCTGCAAC GGTGACTCTG GGGGGCCCCT GATCTGCAAC 720

GGGTACTTGC AGGGCCTTGT GTCTTTCGGA AAAGCCCCGT GTGGCCAAGT TGGCGTGCCA 780

GGTGTCTACA CCAACCTCTG CAAATTCACT GAGTGGATAG AGAAAACCGT CCAGGCCAGT 840

TCTAGACATC ACCATCACCA TCACTAGCGG CCGCTTCCCT TTAGTGAGGG TTAATGCTTC 900

GAGCAGACAT GATAAGATAC ATTGATGAGT TTGGACAAAC CACAACTAGA ATGCAGTGAA 960

AAAAATGCTT TATTTGTGAA ATTTGTGATG CTATTGCTTT ATTTGTAACC ATTATAAGCT 1020

GCAATAAACA AGTTGAC 1037